Structural properties of scale-free networks

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Physik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
der Humboldt-Universität zu Berlin

von
Ramon Xulvi-Brunet

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Christoph Markschies

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:
Prof. Thomas Buckhout, Ph. D.

Gutachter:

1. Prof. Dr. Igor Sokolov
2. Prof. Dr. Bernd Blasius
3. Prof. Dr. Lutz Schimansky-Geier

eingereicht am: 28. April 2006
Tag der mündlichen Prüfung: 22. Juni 2006

Zusammenfassung

Netzwerke sind überall, von der elektrischen Stromversorgung über die Biochemie der Zellen, das Internet bis hin zu sozialen Netzen. Netzwerke als mathematisches Konzept haben sich in den letzten Jahren zu einem wichtigen Werkzeug der Beschreibung komplexer Systeme entwickelt. Ihre grundlegende Eigenschaft ist, dass sie aus einer grössen Anzahl dynamischer Elemente bestehen, die sich gegenseitig beeinflussen und dabei nicht linear gekoppelt sind. Die moderne Netzwerkwissenschaft will die Wechselwirkung zwischen den einzelnen Untereinheiten erklären und davon ausgehend verständlich machen, auf welche Weise Prozesse auf einem Netzwerk stattfinden können. Zum Beispiel wird untersucht, wie die Struktur sozialer Netze die Ausbreitung von Information oder von Krankheiten beeinflusst, wie die Topologie des World Wide Web das Surf-Verhalten oder die Funktionalität von Suchmaschinen beeinträchtigt oder welche Auswirkungen die Hierarchie in ökologischen Nischen auf die Populationsdynamik der einzelnen Spezies hat. Darüber hinaus gilt es herauszufinden, welche grundlegenden Prinzipien der Evolution realer Netzwerke zugrunde liegen, das heißt nach welchen Regeln sich einerseits die Untereinheiten entwickeln und welchen Einfluss andererseits deren Vernetzung hat. Die vorliegende Dissertation beschäftigt sich sowohl mit der Topologie verschiedener Netzwerke als auch mit den der Evolution zugrunde liegenden Prinzipien. Schwerpunkte liegen dabei auf den folgenden zwei Aspekten: erstens dem Einfluss von so gennanten “vertex-pair correlations”, das heißt Korrelationen zwischen den Untereinheiten, auf die Topologie und zweitens der Auswirkung der Geographie auf die Netzwerkentwicklung. Es wird der bedeutende Einfluss aufgezeigt, den die Korrelationen auf wichtige statistische Größen der Netzwerke haben. Weiterhin analysieren wir die Perkolationseigenschaften, die Aufschluss über die Empfindlichkeit gegenüber Störungen in der Vernetzung geben. Damit können zum Beispiel Fragen aus der Epidemiologie diskutiert werden. Es zeigt sich, dass die Topologie vieler Netzwerke und ihre Perkolationseigenschaften deutlich von Korrelationen beeinflusst werden. Schließlich untersuchen wir im letzten Teil dieser Arbeit, wie die Einbettung von Netzwerken in eine endlich-dimensionale Geographie auf die Modellierung und Entwicklung Web-ähnlicher Systeme Einfluss nimmt.

Schlagwörter:
Netzwerke, Graphentheorie, komplexe Systeme, Perkolation

Abstract

Networks are all around us, from electrical power grids to the biochemistry of cells, from the Internet to social webs. The mathematical concept of network has recently been turned into an important tool for describing complex systems, whose principal characteristic is that they consist of a large number of mutually interacting dynamical parts which are coupled in a nonlinear fashion. Modern network science attempts to explain the structure of interactions between the subunits of a system in order to understand their functioning and the processes taking place in them. It tries, for instance, to grasp how the structure of social networks affects the spread of information or human diseases, how the structure of the World Wide Web influences the search engines and surfing behavior, or how the hierarchy of ecological niches affects population dynamics. Beyond this, the ultimate goal of network science is to discover what generating principles exist behind the evolution of real systems. It tries to find the fundamental principles under which the subunits evolve, and the wiring of interactions. This thesis centres both on the study of the topological structure of networks and the analysis of the underlying principles responsible for their evolution. More specifically, it concentrates on the following aspects: the influence of vertex-pair correlations on network topology, the network percolation problem, which is closely related to the spreading of epidemics and the robustness of networks, and the effects of geography as a generating element. We show that important topological and percolation properties change considerably when modifying the connection probabilities between vertices, and that geography as well plays a crucial role in the modeling of evolving real web-like systems.

Keywords:
networks, graph theory, complex systems, percolation

Contents

Chapter 1

Introduction

1.1 Some preliminary words

In the last few years numerous real-world systems have been successfully described by *networks*. The Internet and the World Wide Web are two typical examples of systems which are widely considered man-made *networks*. Other examples of *networked systems* can be found in many different fields, ranging from biology to communication systems: metabolic reactions in the living cell, protein-protein interactions, electrical power grids, ecological food webs, electronic circuits, airline routes, chains of historical events, sexual relations between groups of persons, etc. In effect, networks surround us. More examples are protein folding and genetic regulatory networks, networks in linguistics, scientists' collaboration networks, road and railway networks, neural networks, etc.

A *network* is a collection of entities, which we will call *points*, *nodes* or *vertices*, some of which are joined pairwise. Connections between vertices are usually called *lines*, *links* or *edges*. In this way, the Internet may certainly be modeled as a network. Here the vertices could be the routers and the edges the physical connections between them. The representation of a system by a network is, however, not unique, and depends on the level of abstraction that we use. For instance, the Internet could also be considered a collection of subnetworks, which could be composed of hundreds of routers and computers that are connected together. Both representations are valid and depend on the properties that we want to study in the system and on the level of simplification in our approach. In fact, the Internet has been studied at both presented levels, the router level (where vertices represent routers) and inter-domain or autonomous system level (where subnetworks are vertices). Another well-known network is the World Wide Web. Here, web pages (documents) are the vertices of the network and the edges represent the hyperlinks (URLs) that point from one page to another. Metabolic reactions in the cell are networks where vertices represent chemical substrates (such as ATP, ADP, or H_2O) and edges represent the predominantly directed chemical reactions in which these substrates can participate. Social networks are persons or groups of people (vertices) showing a pattern of contacts or interactions between them (edges). Thus, all systems capable of being modeled as a set of entities which are somehow connected between them can be represented by a network. Given that many physical structures can be described in this way, it is not surprising that networks abound in the world.

Traditionally, the study of networks has been the territory of graph theory, one of the fundamental pillars of discrete mathematics. In 1736 Leonard Euler solved the famous Königsberg bridge problem [Euler, 1736], a problem that played an important role in the development of geometry where distance is not relevant (topology). At the time, Königsberg (a city in old

Prussia, today named Kaliningrad) had seven bridges spanning the various branches of the river Pregel, which divided the town in four parts. The problem under discussion was whether it is possible to take a walk through the town in such a way as to cross over every bridge exactly once. Euler's approach was to regard the four spots of land as points to be visited, and the bridges as paths between those points, as figure 1.1 shows. The problem, put in this way, can be reformulated as whether this corresponding diagram of points and lines, this graph in mathematical terminology, can be traced without tracing any lines more than once. Euler observed that when a point is "visited" in the middle of the process of tracing a graph, there must be a line coming into the point, and another line leaving it; and so the number of lines at the point must be an even number. This must be true for all but at most two of the points: the one you start at, and the one you end at. Thus, a connected graph is traversible if and only if it has a maximum of two points to which an odd number of lines converge. (Note that the starting and ending points may be the same, in which case the number of lines at every point must be even). A quick look at the graph shows that there are more than two points to which an odd number of lines converge, and so the graph cannot be traced. That is, the desired walking tour of Königsberg is impossible. This solution is usually considered as the starting point of the so-called graph theory, our science of networks.

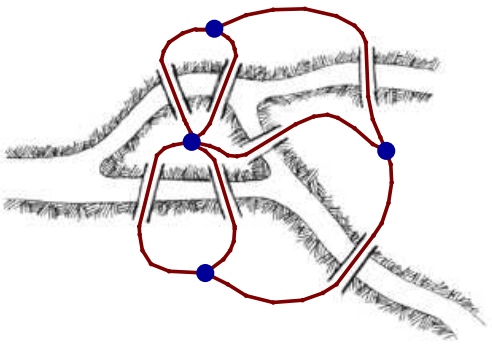

Figure 1.1: The seven bridges of Königsberg. Diagram of the bridges superimpose on a schematic map of Königsberg.

Many optimization questions have turned into classic problems of graph theory. One example is the traveling salesman problem in which, given a number of cities and the cost of traveling from any city to any other city, one must find the cheapest round-trip route that visits each city once and then returns to the starting city. An equivalent formulation in terms of graph theory is: given a *weighted graph*, where the vertices would represent the cities, the edges would represent the roads, and the *weights* would be the cost or distance of that road, find the *Hamiltonian cycle* (closed path in the graph that visits each vertex exactly once) with the least weight (the weight of a path or cycle is the sum of the weights of the traversed edges). Hamiltonian paths and cycles are named after William Rowan Hamilton, who in 1856 invented the mathematical *icosian game*, which involves finding a Hamiltonian cycle in the graph of a dodecahedron [Hamilton, 1931]. A related problem is the bottleneck traveling salesman problem, which consists in finding the Hamiltonian cycle in a weighted graph with the minimal length of the longest edge. These problems, more than a mathematical game, are of considerable practical importance in areas of transportation and logistics.

Another famous question of graph theory is the four color problem. The problem itself dates back to 1852, when Francis Guthrie, while trying to color the map of counties of England, noticed that four colors sufficed. The four color theorem states that every possible geographical map can be colored with a maximum of four colors in such a way that no two adjacent regions receive the same color (two regions are called adjacent if they share a border segment, not just a point). To formally state the theorem, it is easiest to rephrase it in graph theory: every region

of the map is replaced by a vertex of the graph, and two vertices are connected by an edge if and only if the two regions share a border segment. The theorem then states that the vertices of every planar graph (a graph that can be embedded in a plane so that no edges intersect) can be colored with at most four colors so that no two connected vertices receive the same color. It was not until 1977 that the four color conjecture was finally proven [Appel and Hake, 1977a,b]. The proof reduced the infinitude of possible maps to 1936 configurations (later reduced to 1476) which had to be checked one by one by computer. Thus, the four color theorem was the first major theorem to be proven using a computer. (This proof, however, is not accepted by all mathematicians because it can not directly be verified by a human).

While trying to solve these and other problems, mathematicians invented many fundamental graph theoretic terms and concepts. Nevertheless, it was not until the twentieth century that the graph theory was developed into a substantial body of knowledge. Mathematicians, who were up till then interested in optimization problems on more and more complex topological geometries, then started to use methods of probability theory to determine structural properties of large-scale graphs. In the 1950s the Hungarian mathematicians Paul Erdös and Alfréd Rényi exhaustively studied a model with no apparent design principles, usually known in mathematical literature as *random graph* [Bollobas, 2001]. The model, introduced by Solomonoff and Rapoport [1951] and later reinvented by Erdös and Réyi [1959], guided the thinking about networks for decades to come. Erdös and Rényi demonstrated in a series of papers in the 1960s that many structural properties of the random graph are exactly solvable in the limit of large graph size. These results signified the start of modern graph theory [Bollobas, 1998, Harary, 1995].

On the other hand, real networked systems had also been investigated experimentally in the past century. Already in the 1930s sociologists had studied the importance of the patterns of connection between people to the understanding of the functioning of human society. They investigated the *centrality* of the individuals in society (which persons are the best connected to others or have most influence) and their *connectivity* (whether and how individuals are connected to one another through the network). The most popular manifestation of this connectivity is the "six degree of separation" concept, uncovered by the social psychologist Stanley Milgram in 1967, who, by means of ingenious experiments consisting in sending packets between people living in different geographical regions, concluded that there was a path of acquaintances with a typical length of about six between most pairs of people in the United States [Milgram, 1967]. This length is much smaller than we might have guessed, given the number of persons in the United States and the physical distance between people, who also presumably had no direct social or other contacts. This is the idea behind the famous expression *small world effect*, the fact that most pairs of vertices in most networks seem to be connected by a short path through the network.

During the last years, two simultaneous developments have produced a burst of interest in network research. First, the emergence of the complex systems science as a common interdisciplinary subject linking traditional branches like physics, biology, chemistry, or social science. What "complex system" means is not yet well-defined in science, but a common definition is that it is a system consisting of many interacting units whose collective behavior cannot be explained by the behavior of its individual units alone. Since complex system consists of interacting units, networks give the necessary underlying structure to describe them. The second reason for the recent explosion of networks in science is both the computerization of data acquisition in all fields, which has allowed to obtain accurate data of large web-like systems, and the increasing computing power, which has made their investigation feasible, exploring questions that could not be addressed before. Thus, the very human tendency of modeling a complex system by breaking it down to simple interacting components, together with the advent of modern information technology, has created what we could already call a new field in science, the science of

networks. Its roots doubtlessly lie in graph theory, but, unlike this field of theory, which treats only pure topological aspects of graphs, network analysis is oriented to provide new tools capable of depicting the structure of real-world complex systems, and of understanding the processes that take place in them.

1.2 Concepts and measures

One major problem that plagues the network science is the lack of consistency in terminology. Because the nomenclature is not standardized, one must to be careful to check each author's definitions and to specify ones own notation before embarking on talk about the subject. This section is dedicated exclusively to the explication of concepts, measures, and terminology that we will use in the text.

As mentioned, a *graph* or *network*[1] is a collection of dots that may or may not be connected to each other by lines. What dots and lines can represent is not important since we are only interested in the *structure* or *topology* that the connections define on the set of dots. A dot is called a *vertex*[2] and a line is called an *edge*[3]. An edge is defined by the pair of vertices that it connects; the two vertices are called *end-vertices*. The *order* of a graph is the number of its vertices, and the *size* the number of its edges. If the size of a graph is of the same order of magnitude as the order of the graph, then the graph is said to be *sparse*; otherwise it is *dense*. An edge which joins a vertex to itself, that is, which starts and ends on the same vertex, is called a *loop*. An edge is *multiple* if there is at least another edge in the graph which connects the same pair of vertices; otherwise it is *simple*. The *multiplicity of an edge* is the number of multiple edges sharing the same pair of vertices; the *multiplicity of a graph*, the maximum multiplicity of its edges. A graph is a *simple graph* if it has neither multiple edges nor loops. If it has multiple edges but no loops it is called a *multigraph*. A graph which contains both multiple edges and loops is called a *pseudograph*. When stated without any qualification, a graph is almost always assumed to be simple. When an edge is associated with a *weight*, commonly a real non negative number, it is called a *weighted edge*. A graph in which each edge is weighted is called a *weighted graph* or *valuated graph*. In addition, edges may also be endowed with direction. A *directed edge* or *arc* is an ordered pair of end-vertices, the first ones called *head* and the second *tail*; an *undirected edge* disregards any sense of direction and treats both end-vertices interchangeably. A graph is said to be *directed* or that it is a *digraph* if all of its edges are directed; if all of its edges are undirected it is called an *undirected graph* or simply *graph*.

Perhaps the most important concept in graph theory is the notion of *path*. Intuitively, a path is not more than a route that you travel along edges and through vertices in a graph. In modern literature, it is also referred to as *walk* or *open walk*. More precisely, a path is defined as a finite sequence of alternating vertices and edges, beginning and ending always with a vertex, such that each pair of consecutive vertices are connected by the edge which is located between both vertices in the sequence. Two paths with the same starting and ending vertices are *internally disjoint* (some people consider it independent) if they do not have any vertex in common, except the first and last ones. A *simple path* or *elementary path* is a path in which no vertex appears twice. The same ideas can be extended to closed paths. A path having at least one edge, and which begins and ends on the same vertex is called a *closed path*, or *closed walk*, or more often a *cycle* or *circuit*. A circuit or cycle is called *simple* or *elementary* if it does not contain repeated

[1] Mathematicians never use the term "network" but only "graph". In network theory, "network" can refer to both the actual system and the mathematical representation. When we refer to the mathematical representation the terms "network" and "graph" are interchangeable.

[2] Vertices are also called *sites* (physics), *nodes* (computer science), or *actors* (sociology).

[3] Synonyms for edges are *bonds* (physics), *links* (computer science), and *ties* (sociology).

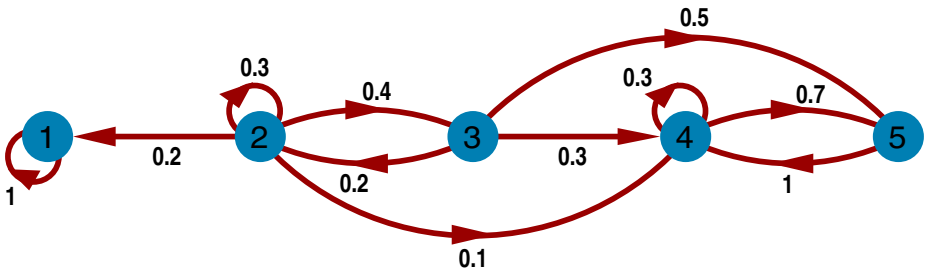

Figure 1.2: Example of a directed weighted pseudograph. In this special case the sum of the weights of all outgoing edges of a vertex is unity. The example corresponds to what is usually called a Markov graph, i. e., a graphical representation of a Markov chain. Vertices represent possible states in which the considered system can be found. Each valued edge represents the transition probability from state i to state j. Although Markov graphs are not "physical" networks, they are a very important application of graph theory in probability theory.

vertices, except that the initial vertex is the terminal vertex[4]. The *length* of a path (closed or not) is simply the number of edges in it. Cycles of length 1 are loops and cycles of length 2 are pairs of multiple edges. All cycles in a simple graph have a length larger or equal to 3. A cycle that has an odd length is called an odd cycle; otherwise it is an even cycle. The *girth* of a graph is the length of a shortest simple cycle in the graph; and the *circumference*, the length of a longest simple cycle. The path between two vertices with the smallest possible length is called a *geodesic*, and its length the *distance* between the two vertices. Note that to define distance between two vertices they must be connected by a path. When the two nodes bear no access to one another, then their distance is defined as infinite. The *eccentricity* of a vertex in a graph is defined as the maximum distance between the vertex and any other vertex. The maximum eccentricity over all vertices in a graph is the *diameter* of the graph, and the minimum, the *radius* of the graph. Trivially, diameter $\leq 2 \cdot$ radius. The eccentricity is a very useful notion to "place" vertices within the graph, or, to put it another way, to define what the "center" and the "periphery" of a graph are: vertices with maximum eccentricity are called *peripheral vertices*, while vertices of minimum eccentricity form the *center*.

In graph theory, degree, especially that of a vertex, is usually a measure of immediate *adjacency*. Two vertices are considered *adjacent* if an edge exists between them. The set of *neighbors* of a vertex, called a *neighborhood* or an *open neighborhood*, consists of all vertices adjacent to the vertex but not including it; when the node is also included, it is called a *closed neighborhood*. The number of edges connected to a vertex is called the *degree*[5] of the vertex. Note that the degree is not necessarily equal to the number of vertices adjacent to a vertex, since there may be more than one edge between any two vertices. A vertex of degree 0 is an *isolated* vertex, and a vertex of degree 1 is called a *leaf*. The *maximum degree* of a graph is the largest degree over all vertices; the *minimum degree*, the smallest. A graph in which every vertex has the same degree is *regular*. It is *k-regular* if every vertex has degree k. In directed graphs one discriminates between *in-degree* (the number of incoming arcs), and *out-degree* (the number of outgoing arcs); and correspondingly the *in-neighborhood* and *out-neighborhood*. If N_k denotes

[4]Here there is much ambiguity in the literature. Sometimes "graph","path", and "cycle" denotes what we have respectively defined as "simple graph","simple path", and "simple cycle". Moreover, the word "loop" is often synonymous to "simple circuit".

[5]To add to the confusion, degree is sometimes referred to as connectivity, a word which already has another meaning in graph theory, as we will see.

the number of vertices of degree k, an ordered list of the N_k values of a particular graph is called a *degree sequence* (*in-degree* or *out-degree sequence* in directed graphs); the degree sequence satisfies $\sum_k N_k = N$. A finite graph of order N, directed or undirected, is often represented by an N-by-N matrix called its *adjacency matrix*, which provides a complete description of the graph. Each element a_{ij} of the adjacency matrix is equal to the number of edges connecting the vertices i and j. *Spectral graph theory* studies relationships between the properties of the graph and its adjacency matrix.

The word "independent" usually carries the connotation of pairwise disjoint or mutually nonadjacent in graph theory. In this way, independence is a form of immediate non-adjacency. A graph can be decomposed into independent sets in the sense that the entire vertex set of the graph can be partitioned into subsets with pairwise disjoint or independent vertices. Such subsets are called *partite sets*, or simply *parts*. Thus, a graph is called *k-partite* if the vertices of the graph can be partitioned into k vertex sets such that no two vertices in the same set are connected by an edge in the graph. For $k = 2$, we have a 2-partite graph, more commonly called a *bipartite graph*. A 3-partite graph is also called a *tripartite graph*. A graph can be proved bipartite if there is no odd cycle. If a graph can be decomposed into k partite sets it is also said to be *k-colorable*. A different notion is related to the word "complete". A *complete graph* is a graph in which all pairs of vertices are adjacent. If N is the number of vertices of a complete simple graph, then it has $N(N-1)/2$ edges. A graph is called a *complete k-partite graph* if it is k-partite, and whenever two vertices are in different *parts* of the graph they are connected by an edge. Finally, a set of isolated vertices is called an *independent set*. Since the graph induced by any independent set is somehow an *empty graph*, the two terms are usually used interchangeably.

Connectivity extends the concept of adjacency and is essentially a form (and measure) of concatenated adjacency. If it is possible to establish a path from any vertex to any other vertex of a graph, the graph is said to be *connected*; otherwise, the graph is *disconnected*. A graph is *totally disconnected* if there is no path connecting any pair of vertices (this is just another name to describe an empty graph or independent set). A subset of vertices is *connected* if a path between every pair of vertices in the subset exists. If the subset cannot be extended without losing their connectedness, it is called a maximal connected subset or *component*. A disconnected graph has two or more distinct components. When a component which is much larger than the other components in the network exists (its order is a finite fraction of vertices, even when $N \to \infty$), then it is called the *giant component* or *supercluster* of the graph. If it is always possible to establish a path from any vertex to every other one, even after removing any $k-1$ vertices, then the graph is said to be *k-connected*. Note that a graph is k-connected if and only if it contains k internally disjoint paths between any two vertices. A *cut vertex*, or *articulation point*, is a vertex whose removal disconnects a graph; the set of cut vertices is called the *cut set* or *separating set*. The *vertex connectivity* or simply *connectivity* is the minimum number of nodes whose deletion from a graph disconnects it. By convention, a complete graph with N vertices has connectivity $N-1$, and a disconnected graph has connectivity 0. An edge whose removal disconnects a graph is called a *cut edge*, or *separating edge*. The *edge connectivity* of a graph is the minimum number of edges whose deletion disconnects the graph. One well-known result of graph theory is that for all graph *vertex connectivity* $\leq$ *edge connectivity* $\leq$ *minimum degree* of the graph. A special type of graphs, in relation to connectivity, are *trees*. A *tree* is not more than a simple, undirected, connected, acyclic graph. A tree with N vertices has $N-1$ edges. Conversely, an undirected connected graph with N vertices and $N-1$ edges is a tree. When all vertices of a tree have the same degree z, then it is usually called a *Cayley tree* (or a *Bethe lattice* by physicists). The deletion of any vertex or edge of a tree disconnects it. When a graph is simple, undirected, acyclic, but disconnected, it is said to be a *forest*. Quite an opposite concept is the notion of *clique*. A *clique* is a complete subset of vertices in the graph, that is, a

subset of vertices in which every pair of vertices are adjacent. A *k-clique* is a clique consisting of k vertices, or to put it another way, a clique of order k. The *clique number* of a graph is the order of a largest clique in the graph.

In the past few years many other new concepts and measures have been proposed and investigated to characterize statistically large-scale networks. Three concepts have commonly occupied a prominent place when it comes to studying real networked systems: the *degree distribution*, the *average shortest path length*, and the *mean clustering coefficient*. The degree distribution, a concept related to the degree sequence of a graph, is defined as the probability function $P(k)$, which gives the probability that a vertex chosen at random in the network has exactly k edges; empirically, $P(k) = N_k/N$, where N_k is the number of vertices of degree k. The *average shortest path length* is the average distance between all vertex pairs of a connected graph[6]. When a graph is disconnected one may talk about the average shortest path length of its components. Moreover, in a disconnected graph in which a giant component exits, it is usually more interesting to find the average shortest path length of this component than the one over the whole graph, which is simply infinity. (The same occurs with the diameter). The clustering coefficient is a quantity which was introduced to measure the local tendency of vertices to form highly connected clusters. To clarify this concept, let us consider the neighborhood of a given vertex i, consisting of k_i vertices. If the neighbors formed a clique (a fully connected cluster) there would be $k_i(k_i-1)/2$ edges between them. The ratio between the number of edges l_i that really exist between the neighbors and the maximum number $k_i(k_i-1)/2$ yields the value of the *clustering coefficient* of the selected node (in a simple graph). The *mean clustering coefficient* of the whole network C is defined as the average of all individual clustering coefficients

$$C = \frac{1}{N}\sum_i \frac{2l_i}{k_i(k_i-1)} \ . \tag{1.1}$$

Averaging C_i over vertices of degree k provides the *degree-dependent clustering function* $C(k)$. Note that all three clustering coefficients are well defined only on undirected simple graphs. Definitions of other coefficients can be found in [Dorogovtsev, 2004, Soffer and Vázquez, 2005]. The property that these other clustering coefficients try to measure is what is usually called *transitivity* or *clustering*, the fact that if vertex i is connected to vertex j and vertex j to vertex l, then there is a heightened probability that vertex i will also be connected to vertex l.

In most kinds of networks there are different types of vertices, and the probabilities of connection between vertices often depends on these types. In corresponding scientific literature this peculiarity is referred to as vertex-pair correlations. For instance, the network of human sexual contacts contains two types of vertices (men and women), and, since most sexual relations are heterosexual, the network has a structure close to a bipartite graph. On the other hand, most social networks show *community structure*, i. e., groups of somewhat similar vertices that have a high density of edges within them, with a lower density of edges between the groups. Another special case of vertex-pair correlations are the so-called degree-degree correlations, which occur when the connections between vertices depend on the degree of vertices. In some networks high-degree vertices associate preferentially with other high-degree vertices (a property referred to as *assortativity*), but in others, high-degree vertices prefer to connect to low-degree ones (*dissortativity*). Vertex-pair correlations have important network structure effects, and we will later analyze them in depth.

[6]Another reason to be cautious with terminology. Several authors use the term "diameter" (the length of the longest geodesic path between any two nodes) to mean "average shortest path length". In addition, this term has many synonyms: *characteristic path length*, *average path length*, *average geodesic distance*, or simply, *average distance*.

1.3 Networks in the real world

The study of networks was initiated by a desire to understand various real systems, ranging from communication networks to biological webs. Thanks to the databases available, nowadays we have copious and accurate information about the Internet, metabolic reactions in the cell, or the so-called coauthor-ship collaboration networks, among others. But, what structural properties have these networked systems? Do all of them present similar features?. Or, on the contrary, are there characteristics in their nature which make them different? In this section we review what we know about the topology of real-world networks.

The Internet, a man-made network designed for the distribution of information, and a very widely studied technological network, is for many common people the archetypal example of what a network is. In this network routers represent nodes and the physical connections between them edges. Faloutsos et al. [1999], Pastor-Satorras et al. [2001] and Vázquez et al. [2002], among others, studied in depth different aspects of the Internet's topology. The collected data demonstrates that the mean number of connections of its vertices noticeably increases as the Internet grows, a phenomenon called accelerated growth. In addition, the degree distribution follows approximately a power law [7], i. e., it falls as

$$P(k) \sim k^{-\gamma} \ , \tag{1.2}$$

for some constant γ (in this case $\gamma \simeq 2.5$). The measured mean clustering coefficient, $C \simeq 0.2$, is much larger than that corresponding to the classical random graph (with a similar number of vertices and edges), $C \simeq 0.001$. Random graphs, because of their randomness (i. e., no apparent design principles) and historical importance, have been turned into the reference in which many topological properties of real-world networks are compared. Investigations about the average path length demonstrate its small-world character.

Other technological networks, like electric power grids [Watts, 1999a, Amaral et al., 2000], airline routes [Amaral et al., 2000], road [Kalapala et al., 2003, Sen et al., 2003] and railway [Latora and Marchiori, 2002] networks, etc, have not yet been sufficiently studied. Electric power grids can be described by networks in which nodes are generators, transformers, or substations, and the edges high-voltage transmission lines. Data on a large power grid were analyzed by Watts and Strogatz [1998]. They showed that the degree distribution of this network is exponential-like. In railway networks, stations are considered as vertices and an arbitrary pair of stations is said to be connected by an edge when at least one train stops at both stations. Sen et al. [2003] studied the Indian railway network and also obtained that $P(k)$ approximately fits to an exponentially decaying distribution. All these networks present very large mean clustering coefficients and, interestingly, values for average shortest path lengths a little larger than most networks.

Classics and well documented web-like systems are the so-called networks of citations between academic papers (see for example de S. Price [1965] and Redner [1998]). Most learned articles cite previous works by others on related topics. These citations form a network in which the vertices are articles and a directed edge from article A to article B indicates that A cited B. Note that in this case article B cannot cite article A, because papers can only cite other papers that have already been written, not those that have yet to be written. Because of this constraint citation networks do not have circuits. How is the number of papers that have been cited k times

[7] Networks whose degree distribution follows approximately a power law are called "scale-free" in network literature. The reason is the following: Scale-free functions are those which obey $f(ax) = bf(x)$ with a and b constants, i. e., those that have the same form after a rescaling process. On the other hand, the only functions which obey this rescaling property are the power law functions. For our purposes, "scale-free" and "power law" will be synonymous.

distributed?. To put it another way, how is their in-degree distributed? The two hitherto studied networks, the database of the Institute of Scientific Information (ISI) for the 1981-1997 period and citations from Physical Review D between 1975 and 1994, show a fat-tailed distribution of the connections, consistent with $P_{in}(k) \propto (k+\text{const})^{-\gamma}$ and $\gamma = 2.9$ [Tsallis and de Albuquerque, 2000]. Recently, Vázquez [2001] has extended these studies to the distribution of the number of references in a paper (the out-degree), finding that it has an exponential tail.

Patterns of friendships between individuals [Rapoport and Horvath, 1961], business relationships between companies [Davis and Greve, 1997], and sexual contacts [Liljeros et al., 2001] are typical examples of connections between people or groups of people in social networks. Traditional social network studies often suffer from problems of inaccuracy, subjectivity, and small sample size. With the exception of a few ingenious indirect studies such as Milgram's [Milgram, 1967], data collection is usually carried out by querying participants directly using questionnaires or interviews. Although much effort is put into eliminating possible sources of inconsistency, it is generally accepted that there are large and essentially uncontrollable errors in most of these studies. Because of these problems many researchers have turned to other methods for probing social networks. One source of copious and relatively reliable data is collaboration networks. Classic examples of such networks are the collaboration network of film actors [Watts and Strogatz, 1998, Newman et al., 2001, Albert and Barabási, 2000] (where actors are the vertices of the network and two actors are considered connected if they have appeared in a film together) and several networks of coauthor-ship among academics [Newman, 2001b,c,d, Barabási et al., 2002] (in which individuals, the vertices, are linked if they have co-authored one or more papers). Histograms of the number of collaborators of scientists are well fitted by a power law form with an exponential cutoff:

$$P(k) \propto k^{-\gamma} e^{-k/\tau} \tag{1.3}$$

where γ and τ are constants. The degree distribution of the other collaboration networks seem to follow curves close to power laws. In addition, all these networks also show large clustering coefficients and small average path lengths.

An important network characterizing the cell describes protein-protein interactions, where the nodes are proteins and they are connected if it has been experimentally demonstrated that they bind together. The study of these physical interactions for *yeast* (a well-studied group of unicellular fungi [YEAST, www]) shows that the degree distribution of their physical protein interaction map follows a power law with an exponential cutoff $P(k) \sim (k + A)^{-\gamma} \exp\{-(k + A)/B\}$ with $A = 1$, $B = 20$, and $\gamma = 2.4$ [Jeong et al., 2001]. Further results about *yeast* can be found in [Ito et al., 2000, 2001]. The distributions of outgoing and incoming edges of the metabolic networks of diverse organisms have been found to follow power laws for all organisms studied, with degree exponents varying between 2.0 and 2.4 [Jeong et al., 2000]; the average path length was found to be approximately the same for all organisms ($l \sim 3.3$. Other metabolic studies focusing on the Escherichia coli bacterium again show a power law degree distribution and small average path length [Wagner and Fell, 2000, Fell and Wagner, 2000]. The clustering coefficient has not been determined because of the directedness of the networks.

A very extensively studied network is the Wold Wide Web. As a citation network it is also a directed network, but circuits are allowed. The Web is characterized by two degree distributions, and both have power law tails: $P_{in}(k) \sim k_{in}^{\gamma}$ and $P_{out}(k) \sim k_{out}^{\gamma}$. Studies on different subsets of the Wold Wide Web [Albert et al., 1999, Kumar et al., 1999, Adamic, 1999, Barabási et al., 2000a, Broder et al., 2000, Kleinberg et al., 1999] have established $\gamma_{in} \simeq 2.1$ and $\gamma_{out} \sim 2.4 - 2.7$. (In spite of the two years delay between the measurements, during which the World Wide Web's order multiplied by five, the measurement $\gamma_{in} = 2.1$ was the same in all studies; the measurements of γ_{out} showed, however, a tendency to increase with time). The World Wide

Web also displays the small-world property: despite the enormous number of nodes ($N \sim 10^9$), the average path length of the Web seems to be smaller than 20 [Broder et al., 2000, Albert et al., 1999, Adamic, 1999]. On the other hand, measurements of clusterization (transitivity) indicate that their mean clustering coefficient, $C \sim 0.1$, is many orders of magnitude higher than the corresponding one of random graphs [Adamic, 1999]. (Because of the directed nature of the Web, which does not allow to use Eq. 1.1 to measure the mean clustering coefficient, the applied approach was to consider each edge bidirectional). Other interesting properties of the World Wide Web can be found in [Flake et al., 2002, Kumar et al., 2000].

It has been observed that, despite functional diversity, most of real web-like systems share similar structural properties. The properties are: power law-tailed degree distributions (that allow the existence of vertices of high degree), small average distance between any two vertices (the so-called small world effect) and a large penchant for creating highly interconnected groups of vertices (vertex clusters with large values of C). These properties are completely different from those of random graphs.

Finally, we point out that many properties of real-world networks have recently been reviewed in some interesting review-articles. Especially complete are [Dorogovtsev and Mendes, 2002, Newman, 2003b, Albert and Barabási, 2002b], but shorter reviews like [Wang and Chen, 2003, Strogatz, 2001, Hayes, 2000a,b] are worthy of reading because of their different but interesting viewpoints. A number of books also make worthwhile reading [Newman et al., 2003, Pastor-Satorras and Vespignani, 2004, Dorogovtsev and Mendes, 2003a, Ben-Naim et al., 2004, Barabási, 2002, Bornholdt and Schuster, 2003, Pastor-Satorras and Rubi, 2003, Watts, 2003]

1.4 Classical models

The graph representation gives an abstract view of the system under study, making it possible to obtain the topological properties of the system by analyzing the representation itself. For small systems, like for example stochastic Markov processes with a small number of states, or Euler Konigsberg's problem, the construction of the graph is not difficult. For large systems, however, collecting data to generate the graph is not an easy task. The construction of the graph is in many cases the subject of research projects requiring large amounts of resources, like the Internet mapping projects [Lumeta, www, CAIDA, www, NLANR, www], the current experiments aiming to determine protein-protein interaction maps [DIP, www, Uetz et al., 2000, Ito et al., 2000, 2001], or the whole research in social science [SW, www]. Such projects usually analyze a representative part of the system, and then extrapolate the findings to the whole network. The difficulty of obtaining accurate data about the global properties (like the average path length or the distribution of circuit lengths), together with the difficulty to obtain analytical results in graph theory, has led the people which work in the field to proceed (in most cases) in a numerical way: first, one constructs a graph using, as a prescription, a few organizing principles, and then one observes the effects of these generating ingredients measuring different features of the generated network.

As physicists, the first type of network that would come to our mind is the regular lattice. Playing a fundamental role in solids, they are characterized by invariance under translation by one lattice spacing along a lattice axis. In this case, the vertices of the network could be the atoms of the crystal and the edges could indicate the most important interactions. For instance, in a simple two dimensional regular square lattice 4.1, each vertex is attached to four edges so that the degree of every vertex is four. An essential property of a regular lattice is that it is always connected. This seems obvious, but is not a triviality in network science. Deciding if a graph is connected or disconnected is one of the first things one should look at. Thus,

studying the connectedness of random graphs was part of the seminal work of Erdös and Rényi. The second characteristic property is that the diameter of a D-dimensional lattice will grow as $N^{1/D}$. For most networks it grows much slower as $\ln(N)$. Thus, in a large random network every vertex is much closer to all the other vertices than in a comparably sized lattice (with the same number of vertices and edges).

The first significant model which appeared in technical literature was the *random graph*. The model is extremely simple: take some number N of vertices and connect each pair with probability p. Consequently, the total number of edges, L, is a random variable with the expectation value $\langle L\rangle = pN(N-1)/2$. An alternative construction consists of taking an initially empty graph with N vertices and then adding L edges one by one so that each edge connects randomly chosen pairs of vertices. The graphs obtained at different stages correspond to larger and larger connection probabilities p; both constructions are equivalent in the thermodynamic limit, i. e., when $N \to \infty$. The greatest discovery of Erdös and Rényi was that the random graph undergoes what we would now call a phase transition: for large N, a giant component appears when the number of edges is $L \geq N/2$, and the graph is likely to be connected only if at least $N\ln(N)/2$ edges are present in the network [Bollobas, 2001, Erdös and Rényi, 1960, Erdös and Réyi, 1961, Cohen, 1988]. Put another way, if one increases the number of edges added randomly to the graph, then there is a sudden change in the connectedness of the network.

In the random graph the number of vertices having exactly k edges is distributed according to a binomial distribution [Bollobás, 1981, Erdös and Réyi, 1959],

$$P(k) = \left(\begin{array}{c} N-1 \\ k \end{array} \right) p^k (1-p)^{N-1-k} \tag{1.4}$$

which in the limit of large N can be approximated to the Poisson form $P(k) \simeq e^{-\langle k\rangle}\langle k\rangle^k/k!$, with $\langle k\rangle \simeq pN$. The typical appearance of the degree distribution, $P(k)$, of the random graph is shown in figure 1.3. The mean clustering coefficient can easily be calculated for this model. If we consider a node in a random graph and its nearest neighbors, the probability that two of these neighbors are connected is equal to the probability that two randomly selected nodes are connected. Thus, the clustering coefficient of a random graph is $C = p \simeq \langle k\rangle/N$. The diameter of a random graph has been studied by many authors [Klee and Larman, 1981, Bollobás, 1990, Chung and Lu, 2001]. Perhaps the most important conclusion is that the diameter of the giant cluster is proportional to $\ln(N)$. As for the average path length, simulations indicate that it also scales logarithmically with the number of vertices N.

The celebrated *small-world model* meant the first successful attempt to generate graphs with a high degree of clustering and, at the same time, with small average distances. The model was also the first model that took the geographical component of real systems into account; the vertices in most real networks are positioned in space, and the geographical proximity presumably plays a role in deciding which vertices are connected to which others. This model has its roots in social systems, in which most friends and acquaintances of a person are their immediate geographical neighbors, and only one or two friends live in other countries or are simply a long distance away. The small-world model starts from a low-dimensional regular lattice in which a low density of large connections that join remote parts of the lattice to one another can be created by adding or moving edges at random. The underlying idea is to combine the large degree of clustering of a regular lattice with the small-world effect caused by the existence of the long-range connections. The rewiring or moving of edges produces a drop both in C and l. By varying the density of long-range edges one can interpolate between a fully ordered lattice and a regular random graph, and in this way obtain low path lengths and at the same time high transitivities.

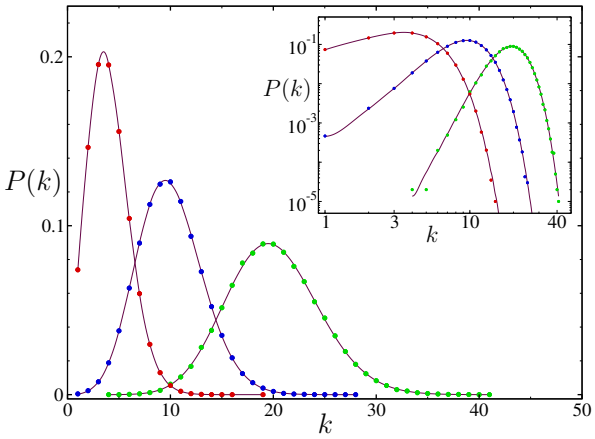

Figure 1.3: The degree distributions, $P(k)$, that result from three numerical simulations of a random graph: $L = 2 \cdot 10^5$ (red points), $L = 5 \cdot 10^5$ (blue points), and $L = 10^6$ (green points). For all three $N = 10^5$. The plot compares $P(k)$ with the corresponding values of the Poisson distribution (curves). The inset shows the same on double logarithmic scales (to compare better with later results of scale-free networks).

Two versions of the small-world model exist. The original model, which was proposed by Watts and Strogatz [1998], Watts [1999a,b], starts from a one-dimensional lattice of N vertices with periodic boundary conditions (a ring), and join each vertex to its k nearest neighbors ($k/2$ on either side). In order to have a sparse but fully connected network at all times it is considered $N \gg k \gg 1$. The system then has $L = Nk/2$ edges. The long-range connections are created by taking a small fraction of edges, p, and rewiring them such that loops and multiedges are excluded. The rewiring procedure involves moving one end of each chosen edge to a new location selected at random from the lattice; if this rewiring procedure produces multiple or self-connections, the step is discarded. By adjusting p, one can closely vary the transition between order ($p = 0$) and randomness ($p = 1$). Watts and Strogatz showed by means of numerical simulations that an interval of values of p exists for which the graph has a large degree of clustering and low average distance. In the variant proposed by [Newman and Watts, 1999a, Monasson, 1999], instead of the edge rewiring procedure, a fraction of new edges joining randomly chosen vertex pairs are added to the lattice. The density of introduced edges is governed by the new parameter (also called) p, that now is defined in such a way that the number of introduced edges is Lp and, therefore, the mean degree is $\langle k \rangle = 2L(1 + p)$. The variant is easier to analyze than the original Watts-Strogatz model since it has the property that the graph always remains connected. Both versions of the small-world model have been studied in some detail in scientific literature [Barrat and Weigt, 2000, de Menezes et al., 2000, Newman, 2000, Newman et al., 2000, Barthélémy and Amaral, 1999a,b, Farkas et al., 2001, Pandit and Amritkar, 1999].

However, the small-world model does not generate networks with power law tailed degree distribution. This problem is attacked by models of network growth. This type of model arises as an attempt to answer the question of what mechanisms are able to explain the features of real networks. The question passed from how a graph with the correct topological properties can be generated to what dynamic principles are responsible for the emergence of the scale-free networks. The archetypal model was proposed by de S. Price [1976] (the same de S. Price who already in 1965 studied for first time the in- and out-degrees of citation network), who was inspired by the thought-provoking studies of Simon [1955]. Both already made use of the assumption which is now widely accepted as the probable explanation for the power law degree distribution, the concept of "the rich get richer" or, in Price's terminology, the "cumulative advantage" (see for more details the excellent review Newman [2003b], and also Bornholdt and Ebel [2001]). The idea is that the amount you get goes up with the amount you already have. Unfortunately, the model practically went unnoticed in the scientific community. The concept of cumulative advantage was recently reinvented by Barabási and Albert under the

name *preferential attachment*. In a very influential article [Barabási and Albert, 1999] proposed a simple model based on a few organizing mechanisms capable of reproducing the observed stationary scale-free distributions of many real networks. The work definitively established that the development of large networks is governed by robust self-organizing phenomena that go beyond the particularities of the individual systems. This new viewpoint has inspired a number of modifications and generalizations of the model, so much that the article is considered as the driving force behind most of the recent investigations in the field.

The two basic ingredients of the Barabási and Albert [1999] construction are *growth* and *preferential attachment*. *i*) *Growth*: It is inspired by the fact that most real networks form by the continuous addition of new vertices to the system, so that the number of vertices, N, increases throughout the lifetime of the network. The model starts with a small number, m_o, of vertices to which, at every time step, we add a new vertex with m ($m \leq m_o$) edges that link the new vertex to m different vertices already present in the system. *ii*) *Preferential attachment*: It is assumed that the probability Π that a new vertex connects to vertex i depends on the degree k_i of that vertex, such that

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \ . \tag{1.5}$$

After t time steps this procedure results in a network with $N = m_o + t$ vertices and mt edges. Barabási and Albert demonstrated that this algorithm creates a network whose degree distribution approximately falls following the power law function $P(k) \sim k^{-3}$ using a continuum approach which is based on the assumption that all vertices of the same age have the same degree [Barabási and Albert, 1999, Barabási et al., 1999].

The model of Barabási and Albert can be solved exactly in the thermodynamic limit, $N \to \infty$ (see next chapter). Using the master-equation approach, Dorogovtsev et al. [2000] and, independently, Krapivsky et al. [2000] showed that the probability that a vertex has k edges verifies

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \ , \tag{1.6}$$

which gives the above result $P(k) \sim k^{-3}$ in the limit of large k (see also Bollobás et al. [2001], Krapivsky and Redner [2001], Krapivsky et al. [2001], Kullmann and Kertész [2001]). A deep study of the model shows that it exhibits two types of correlations: one type is the correlation between the age of vertices and their degrees, and the another type are the so-called degree-degree correlations. The first one evidences that the older a vertex the higher is its degree (on average). This result does not coincide with the features of some real networks, among others, with those of the World Wide Web, one of the systems for which the model was initially conceived. For a discussion of the problem see [Adamic and Huberman, 2000, Barabási et al., 2000b, Bianconi and Barabási, 2001a]. The second type deals with correlations between the degrees of adjacent vertices. Krapivsky and Redner [2001] proved that these correlations develop spontaneously as a result of the dynamical process. The weak point of the model is the degree of clustering. Numerical simulations show that the generated networks still have a very small C (in comparison to real networks), which decreases with N following approximately a power law $C \sim N^{0.75}$ [Albert and Barabási, 2002b]. The average minimum path distance grows as $\ln(N)$ for $m = 1$ and logarithmically with N, but with a double logarithmic correction, $\ln(N)/(\ln(\ln(N)))$, for $m \geq 2$ [Bollobás and Riordan, 2004]. The spectral properties of the model have also been studied by [Farkas et al., 2001, Goh et al., 2001].

Chapter 2

Statistical properties of evolving networks

The Barabási-Albert construction captures two basic mechanisms capable of generating power law degree distributions: growth and preferential attachment. The model, however, still displays manifest limitations when it comes to reproduce the exact features of real networks. For example, it predicts a power law degree distribution with a fixed exponent $\gamma = 3$, while, on the one hand, the exponents measured for real networks vary (principally) between 1 and 2 for biological networks and 2 and 3 for non-biological, and, on the other hand, the degree distribution of many real networks have non-power law features (such as exponential cutoffs), which it cannot explain. The degree-degree correlations that the model spontaneously generates are not in perfect agreement with those found in most real networks. The model deals additionally with undirected networks, whereas many real networks are directed. One should remember that it generates networks with very a small degree of clustering, while real networks typically present quite large values of C. Thus, the question that now has to be addressed is if extensions and other variations of the model can make it a more realistic representation of processes taking place in real-world networks. Taken together, if other organizing principles can capture more exactly the observed characteristics. We will describe a few of these extensions in this chapter. Two of them will be discussed in the last section in detail since they will be extensively used in following developments.

2.1 Beyond the Barabási-Albert construction

A crucial element of the Barabási-Albert construction is the assumption of linear preferential attachment. By studying the time evolution of different networks, like the Internet and several collaboration and citation networks, Jeong et al. [2003] and Newman [2001a] conclude that this assumption is well supported by empirical results. They measure the number of new edges that each vertex acquires in a time interval ΔT, and then compare it with the previously existing degree of the vertex. In this way they show that in each case $\Pi(k_i) \sim k_i^{\alpha}$, with the scaling exponent $\alpha \simeq 1$ (for the most networks) or $\alpha \simeq 0.8$ (for a few networks like the neuroscience co-authorship and the actor collaboration networks). Krapivsky et al. [2000] and Krapivsky and Redner [2001] go beyond this and consider the case where the probability of attachment $\Pi(k)$ depends on a general power of k, $\Pi(k) \propto k^{\beta}$. This interesting model, which can be solved using the master-equation approach, exhibits three distinct classes of behavior depending on the value of β. In the sub-linear regime ($\beta < 1$) the degree distribution is given by a stretched exponential, which may also be approximated by a power law multiplicate by an exponential like Eq. 1.3. For

$\beta = 1$, the authors evidently recover the linear preferential attachment and power law degree distributions. For the supra-linear regime ($\beta > 1$) there is no analytical solution. However, it can be deduced that a single vertex gets a finite fraction of all edges in the thermodynamic limit. Moreover, if $\beta > 2$ all vertices will be connected to this "winner" vertex, and the degree distribution (excluding the "winner") will decay exponentially.

The World Wide Web [Broder et al., 2000], the Internet [Faloutsos et al., 1999], and the co-authorship network of scientists [Barabási et al., 2002], among others, show an increase of their average degree $\langle k \rangle$ with time. This phenomenon, in which the number of edges increases faster than the number of vertices, is called *accelerated growth*. A number of models of accelerated growing networks have been proposed in the last few years [Dorogovtsev and Mendes, 2001b, 2003b, Sen, 2004, Gagen and Mattick, 2005]. Let us here consider one of them. Dorogovtsev and Mendes [2001b] analyze the distribution of incoming edges of the following directed network: At every step a new vertex is added to the network, which receives m incoming edges from some non-specified (randomly selected) old vertices. Additionally, $c_o \cdot t^{\alpha}$ (c_o and *alpha* are constants) new directed edges are distributed between the old vertices in such a way that each new edge goes out from a non-specified vertex and is directed to a vertex i chosen with a probability proportional to the sum of its in-degree and a constant A (initial attractiveness). The authors demonstrate that this model has a power law in-degree distribution $P(k) \propto k^{-\gamma(\alpha)}$ controlled by the parameter α, with $\gamma = 1 + 1/(1 + \alpha)$. Thus, the suggested process offers a possible mechanism by which the exponent of the degree distribution can be tuned. The recent work of Gagen and Mattick [2005] shows, however, that some very rapidly accelerating networks display a transition from scale-free to exponential statistics with network growth[1]. On the other hand, Barabási et al. [2002] propose another model in which $\langle k \rangle$ increases linearly in time, in agreement with some measurements on the real co-autor network.

Other extensions of the Barabási-Albert model take into account the finite life-time of vertices, or their finite capacity to connect to other vertices. Thus, Amaral et al. [2000] study an evolving model also based on the ingredients of growth and preferential attachment, but where no new edges may connect to an old vertex if it has already reached a certain age or a critical degree. Numerical simulations indicate that the degree distribution preserves the scale-free character for small k, while an exponential cutoff appears for large k. Dorogovtsev and Mendes [2000a] suggest another variation which makes sense for citation networks. They, inspired by the fact that old papers are rarely cited, investigate another model in which a new vertex is connected to an old vertex with probability proportional (i) to the connectivity of the old vertex (linear preferential attachment) and (ii) to $(t - t_i)^{-\alpha}$, where $t - t_i$ is the age of vertex i (gradual aging). The authors show that the degree distribution of this model depends dramatically on the parameter α: when $\alpha < 1$ the degree distribution follows a power law function whose scaling exponent depends strongly on α, while a function decreasing more rapidly than a power law appears when $\alpha > 1$. More recently the impact of aging has been studied, not only on the degree distribution of networks, but also on the transitivity and degree-degree correlations [Zhu et al., 2003, Hajra and Sen, 2004]. Here, it is concluded that aging can strongly influence network structure, although the more significant effects appear when the intensity of aging is large. On the other hand, Hwang et al. [2005] and Daido and Nakanishi [2004] have obtained interesting results when studying synchronization processes on aging networks.

One important feature of the original Barabási-Albert construction is that the oldest vertices have the highest degrees. As above discussed, the World Wide Web does not exhibit such a correlation between age and degree. Adamic and Huberman [2000] suggest that this is because the probability Π depends also on the intrinsic worth of vertices (for example, a combination of

[1] They argue that such a transition explains, for example, the evolutionary record of single-celled organisms.

interesting content and marketing in the case of the World Wide Web). Bianconi and Barabási [2001a] propose a model in which each new vertex i is given a "fitness" η_i that represents its attractiveness or ability to compete for edges with other vertices. Fitness values are selected from some distribution $\rho(\eta)$, and the new vertex i connects to old vertices according to $\Pi(k_i, \eta_i) \propto k_i\eta_i$. The form of the degree distribution $P(k)$ depends obviously on the form of $\rho(\eta)$. The many variations on the fitness theme basically show that for reasonable selections of the fitness distribution the scale-free behavior is preserved, although the exponent is usually affected by the chosen $\rho(\eta)$ [Ergün and Rodgers, 2002, Calderelli et al., 2002, Krapivsky et al., 2002, Bianconi and Barabási, 2001b, Servedio et al., 2004, Zheng et al., 2003, Lee et al., 2004, Sotolongo-Costa and Rodgers, 2003].

In real systems a number of mechanisms exists for network growth. Not only the mechanism of addition of vertices to the system exists, but also other ones like addition or rewiring of edges or removal of vertices or edges. One example is the World Wide Web, where new links are frequently added between old pages, or old links are rewired or removed. Dorogovtsev and Mendes [2000b] have studied interesting cases. They suggest the following two simple models in which only one extra mechanism is added to the standard Barabási-Albert prescription: At every time step, some new edges (model a) are introduced between unconnected pairs of old vertices i and j with a probability proportional to the product of their degrees k_ik_j, and some existing edges (model b) between the old vertices are removed with equal probability. The exact form of the degree distribution depends on the amount of added or removed edges, but interestingly their power law character is maintained as long as these amounts are not extremely large. Another model, which combines addition and rewiring of edges, is discussed by Albert and Barabási [2000]. The model starts with m_o isolated vertices, and at each time step one of three possible operations is performed: (i) With probability p, m ($m < m_o$) new edges are added to the system, one end-vertex being randomly chosen and the other taken with probability $P(k_i) \propto k_i + 1$. (ii) With probability q, m old edges are rewired; for each rewiring process, a vertex i of the network is randomly chosen, then one of the edges attached to vertex i (say l_{ij}) is selected at random, and finally, this edge is removed and replaced with another edge, l_{in}, where the end-vertex n is again chosen with probability $\Pi(k_n) \propto k_n + 1$. (iii) With probability $1 - p - q$ a new m-degree vertex is introduced, and their m edges connect with probability Π to vertices already present in the system. The network structure of this model is evidently quite complex, and depends strongly on the probabilities p and q. However, it is possible to show that a generalized power law $P(k) \propto [k + f(p, q, m)]^{\gamma(p,q,m)}$ develops over a wide range of values of the parameters. Other similar models also indicate both power law and exponential degree distributions depending on the model parameters [Krapivsky et al., 2002, Sotolongo-Costa and Rodgers, 2003, Tadić, 2001, 2002].

Some authors have imagined alternative mechanisms to the preferential attachment in order to generate power law distributions. In fact, there are networks for which preferential attachment does not seem to be the appropriate model to explain their scale-free character. Examples are protein interaction networks, for which there is no reason to suppose that they evolve according to the preferential attachment prescription [Wagner and Fell, 2000, Jeong et al., 2001, Stelling et al., 2002, Fell and Wagner, 2000, Jeong et al., 2000], and the World Wide Web, for which a possible growing mechanism could be the fact that new Web pages are dedicated to a certain topic link from already existing pages on the same topic [Kleinberg et al., 1999, Kumar et al., 2000]. They propose an interesting growth mechanism for the World Wide Web usually called the *vertex copying mechanism*. We will, however, concentrate here on another very similar model, which seems to be suitable for biological networks, the *duplication model*. Basically it is as follows: At each time step, an old vertex, u, of the network is randomly selected. Then, a new vertex, v, is added to the system in such a way that each of its outgoing edges attaches to a non-

specified vertex of the network with probability p, and to a neighbor, w, of u with probability $1-p$. Chung et al. [2003] and Bhan et al. [2002] prove that this model does generate power law degree distributions. Moreover, the model is able to generate power law graphs with a scaling exponent $\gamma < 2$, as usually seen in biological networks. In a more general form, the duplication model also incorporates mechanisms for the removal of edges [Kim et al., 2002, Vázquez et al., 2003]. Models of protein networks that make use of other mechanisms have been proposed by a number of authors [Eigen and Schuster, 1979, Berg et al., 2004, Wagner, 2003].

The most of the above extensions focus on the behavior of the degree distribution, but other quantities like length distribution of circuits [Bianconi and Capocci, 2003, Bianconi et al., 2005] or degree-degree correlations [Boguñá et al., 2003, Barrat and Pastor-Satorras, 2005], are also essential for characterizing the network structure. A number of new models aimed to reproduce other features of networks have appeared recently. Catanzaro et al. [2004] have proposed a model for scientific collaboration networks capable of mimicking the *assortative* behavior (tendency of high degree vertices to connect to each other) exhibited by many social networks [Newman, 2002, 2003a]. Inspired by the preferential attachment rule, the main variation of this model consists again in allowing growth by addition of new edges between old vertices: With probability p a new vertex is connected to old vertices like in the Barabási-Albert construction. With probability $1-p$ a new edge is added between two unconnected old vertices; the first end-vertex (for example with degree k_1) is determined again by preferential attachment, while the second is selected according to the conditioned probability $1/(|k_1 - k_2| + 1)$. This model shows a power law degree distribution with a tunable scaling exponent and, at the same time, assortative mixing. Gómez-Gardeñes and Moreno [2004] suggest another model based on certain "local" rules which, apart from being able to generate scale-free networks, shows degree-degree correlations and values of clustering close to those measured for real-world networks.

2.2 Properties of important scale-free constructions

In this section we concentrate on the study of two important models: we will call them the *intrinsic-worth* model and the *exact scale-free* construction. The first one differs from the Barabási-Albert construction only in one aspect, namely, that vertices are endowed with a certain intrinsic *worth* or *attractiveness*. Quantitatively, this new ingredient is implemented by means of a small change in the preferential attachment: the probability that the new vertex n connects to vertex i is proportional to $k_i + A$, where k_i is the degree of vertex i and A is a new constant whose role is to introduce the intrinsic worth or *initial attractiveness*. The model reduces evidently to the Barabási-Albert construction when $A = 0$. Note that if $A = 0$, no isolated vertex can ever increase its degree. (In fact, the Barabási-Albert construction does not take into account the real-world possibility that isolated vertices may be "discovered" and linked to the system.) The second model we consider, the *exact scale-free* construction, does not start from any physical generating principles, like the above models. It is a purely mathematical algorithm conceived to construct networks with the desired degree distribution $P(k) = \text{constant} \cdot k^{-\gamma}$. We study both models in some detail here, since most of our later numerical results are based on these two types of networks.

We begin with the *intrinsic-worth* generalization. This model has been examined in depth by Dorogovtsev et al. [2000] and Krapivsky and Redner [2001]. We implement the model as follows: We start with an even number m_0 of vertices and bind them in pairs with one edge each. This will be our initial condition. As in the Barabási-Albert model, at every time we add step a new vertex to the system (linear growth), and connect it by means of m edges ($m \leq m_0$) to m different vertices already present in the system. After t time steps the algorithm results

in a network with $t + m_0$ vertices and $mt + m_0/2$ edges. In contrast with the Barabási-Albert model, the probability Π for the new vertex n to be connected to an old one i depends not only on the degree k_i of vertex i, but also on a constant A:

$$\Pi(k_i) = \frac{k_i + A}{\sum_j (k_j + A)} \ . \tag{2.1}$$

Here, the sum in the denominator includes all vertices in the system except the newly introduced one, and A is a real parameter describing the minimum worth of vertices. For $A \to \infty$, the effects of preferential attachment disappear, and all vertices have the same probability to be chosen by the new vertex; on the other hand, for $A \to 0$ the preferential attachments recover their importance, and the structure of networks tends to that of the original Barabási-Albert construction. (Note that our initial condition is slightly different from that of Barabási and Albert, where the initial m_0 vertices are not connected: in our case all vertices introduced at $t = 0$ have exactly one edge, which allows us to use Eq. 2.1 from the very beginning, even when $A = 0$. This simplifies the algorithm, since we do not have to distinguish between the initial and the further steps. The only difference with the genuine Barabási-Albert construction is that at time t one has $mt + m_0/2$ instead of mt edges present. This initial condition does not influence the asymptotic behavior of the model $t \to \infty$. We only consider non-negative values of A, however, negative values $A > -m$ may also be considered [Dorogovtsev et al., 2000, Krapivsky and Redner, 2001].)

The degree distribution of this model can be analytically obtained using the *master-equation* approach. Following [Dorogovtsev et al., 2000], we study the probability $p(k, s, t)$ that at time t vertex s has degree k. The master equation governing this probability $p(k, s, t)$ is:

$$p(k, s, t+1) = \frac{m(k-1+A)}{\sum_j (k_j + A)} p(k-1, s, t) + \left(1 - \frac{m(k+A)}{\sum_j (k_j + A)}\right) p(k, s, t) \tag{2.2}$$

where the denominators extend over all vertices in the network at time t. We denote N the network's order at time t (or the number of old vertices at time $t + 1$, that is, excluding the new vertex), and N' as the network's order at time $t + 1$ (including the new vertex). Hence, it is verified that $\sum_j k_j + A = 2L + NA = (2m + A)t + (1 + A)m_o$, where $L = mt + m_o/2$ is the number of old edges, and $N = mt + m_o$ the number of old vertices. The first term of Eq. 2.2 corresponds to the probability that vertex s receives one edge of the new vertex assuming that at time t it has degree $k - 1$, while the second term corresponds to the probability that vertex s receives no edge assuming that at time t it already has degree k. Of course, the new vertex, which is introduced at $t + 1$, has degree m. The degree distribution at time t can then be obtained adding the probabilities $p(k, s, t)$ for all vertices of the network divided by N

$$P(k, t) = (1/N) \sum_s p(k, s, t) \ . \tag{2.3}$$

We now evaluate the sum $\sum_s$ in the two terms of the Eq. 2.2. The sum goes over every vertex s except for the new ones. We obtain:

$$\sum_s^N p(k, s, t+1) = \sum_s^N \left[\frac{m(k-1+A)}{2L + NA} p(k-1, s, t) + \left(1 - \frac{m(k+A)}{2L + NA}\right) p(k, s, t) \right] \ . \tag{2.4}$$

Taking into account that $\sum_s^N p(k, s, t+1)$ (where the newly introduced vertex has been excluded) holds $\sum_s^N p(k, s, t+1) = \sum_s^{N'} p(k, s, t+1) - \delta_{kn} = N' P(k, t+1) - \delta_{kn}$, we can rewrite Eq. 2.4

as follows:

$$N'P(k,t+1) - NP(k,t) = \frac{mN}{2L+NA}\left[(k-1+A)P(k-1,t) - (k+A)P(k,t)\right] + \delta_{km} \quad (2.5)$$

Note that this equation is exactly the one proposed by Krapivsky and Redner [2001] in their rate-equation approach. If we now substitute $N = mt + m_o$ and $L = mt + m_o/2$ in Eq. 2.5 and take into account that $P(k,t+1) \simeq P(k,t)$ when $N \to \infty$, we obtain the following expression:

$$P(k,t)\left[1 + \frac{m(t+1+m_o)(k+A)}{(2m+A)t + (1+A)m_o}\right] = \left[\frac{m(t+1+m_o)(k-1+A)}{(2m+A)t + (1+A)m_o}\right] P(k-1,t) + \delta_{km} \quad (2.6)$$

Note that for $t \gg \max\{m_o+1, (1+A)m_o\} \simeq 10$ Eq. 2.6 ceases to depend on t, which means that the degree distribution reaches the stationary state very quickly ($N \sim 10^3$). In the thermodynamic limit, Eq. 2.6 may be approximated by

$$\begin{cases} P(k) = \left[\dfrac{m(k-1+A)}{2m+A+m(k+A)}\right] P(k-1) & \text{if} \quad k > m \\ P(k) = \dfrac{2m+A}{2m+A+m(k+A)} & \text{if} \quad k = m \end{cases} \quad (2.7)$$

This last equation can be easily solved in an iterative way. The solution, which can be expressed as product of the terms of a sequence, or, in a more compact way, when making use of the *gamma* function Γ, is as follows:

$$P(k) = \prod_{i=m+1}^{k} \frac{A+i-1}{A+A/m+i+2} P(m) = \left(2 + \frac{A}{m}\right) \frac{\Gamma(k+A)\Gamma(m+A+A/m+2)}{\Gamma(m+A)\Gamma(k+A+A/m+3)} \; . \quad (2.8)$$

The resulting degree distribution of such networks then tends to $P(k) \sim k^{-3-A/m}$ for $k \gg 1$, which proves their asymptotic scale-free character. Note that the degree distribution of this model (Eq. 2.8) reduces exactly to the degree distribution of the Barabási-Albert model (Eq. 1.6) when $A = 0$.

Let us consider the following example: $A = 4$ and $m = 2$. In this case the degree distribution (Eq. 2.8) reduces to the more simply expression

$$P(k) = \frac{12096}{(k+4)(k+5)(k+6)(k+7)(k+8)} \; , \quad (2.9)$$

from which it is easy to calculate the first moments of the distribution: $\langle k \rangle = \sum_{k=m}^{\infty} kP(k) = 4$ and $\langle k^2 \rangle = \sum_{k=m}^{\infty} k^2 P(k) = 28$. The first moment of $P(k)$ does not in fact yield much information; it confirms only that the size of these networks is m times their number of vertices ($\langle k \rangle = 2L/N$). The fact that $\langle k^2 \rangle$ is finite however does make this case very interesting, as the finiteness or infiniteness of the second moment $\langle k^2 \rangle$ plays an important role on the percolation of networks (see chapter 4). Note that the Barabási-Albert construction ($A = 0$) has $\langle k^2 \rangle = \infty$. This latter can be easily proven; in this case $\langle k^2 \rangle = 2m(m+1)[\sum_{k=m}^{\infty} k/((k+1)(k+2))] = 2m(m+1)[\sum_{k=m}^{\infty}(2/(k+2) - 1/(k+1))] = 2m(m+1)[\sum_{k=1}^{\infty} 1/k - \sum_{k=1}^{m+1} 1/k - 1/(m+1)]$, which is evidently ∞ because the first sum $\sum_{k=1}^{\infty} 1/k$ diverges and all the other terms are finite. In chapter 4 we will investigate the percolation properties of both examples, $A = 0$ and $A = 4$.

Figure 2.2 shows the degree distribution $P(k)$ of this model on double logarithm scales. Panel (a) shows $P(k)$ as a function of k for the preceding example, $A = 4$ and $m = 2$; panel (b) shows also $P(k)$ for the parameter values $A = 0$ and $m = 2$ of the same model (the Barabási-Albert

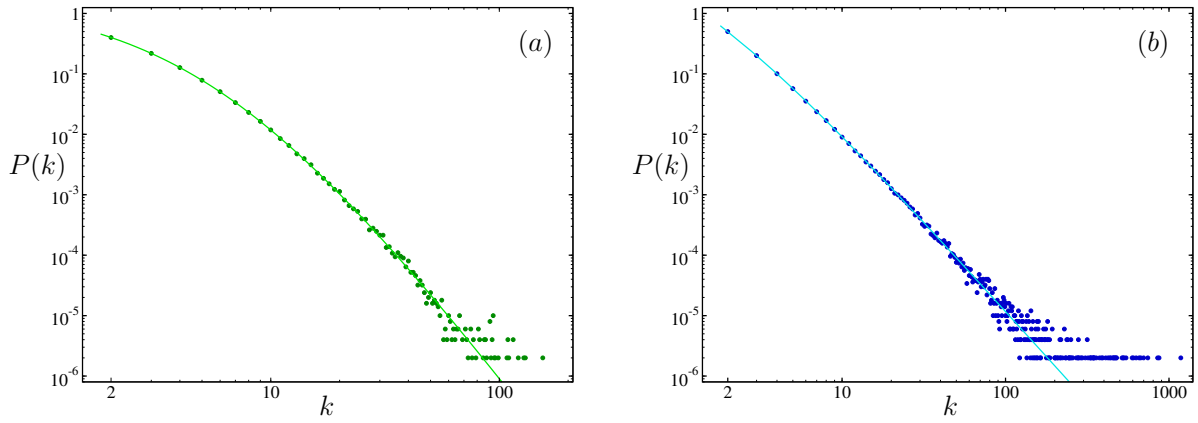

Figure 2.1: Degree distribution $P(k)$ of the *intrinsic-worth* model *versus* k. Panel (a): $A = 4$, $m = 2$; panel (b): $A = 0$, $m = 2$ (Barabási-Albert). The points correspond to numerical results and the curves to the theoretical expression, Eq. 2.8. See text for more details.

construction). In both examples the generated networks have 10^5 vertices, and the numerical results are averaged over five realizations of the model. The points of the graphs are the results of simulations, while the curves correspond to the theory, Eq. 2.8. Notice the perfect agreement between the theoretical degree distribution $P(k)$ of the model, Eq. 2.8, and the numerical results.

The second construction is the *exact scale-free* (ESF) *model*. The aim here is to construct networks of N vertices whose degree distribution decays exactly as $P(k) = ak^{\gamma}$, where a is the normalization constant of the degree distribution. To this end, we propose an algorithm to build networks with a given degree distribution, or a given degree sequence. We implement it in the following way: (i) Create a set of N vertices, indexed by $i = 1, 2, ..., N$. (ii) Assign to each vertex i an integer k_i (its degree) from the probability distribution $P(k) = ak^{-\gamma}$. (Note that the assignment could also be drawn from a desired degree sequence or another arbitrary degree distribution). The assignment must be subject to the constraints $m \leq k < N$ and $\sum_i k_i$ even. These constraints are imposed because no vertex can have a degree larger than $N-1$ in a simple graph, and because we are usually interested in networks where no vertices have a degree smaller than a certain m, the minimum degree in the network. The reason for the last condition, $\sum_i k_i$ even, will become clear when reading the next step of the algorithm. (In this case in which $\sum_i k_i$ is odd, we lower the degree from the largest degree vertex by one). (iii) Select a pair of vertices, i and j, and place between them an edge, l_{ij}, if: $i \neq j$ -so as to avoid the appearance of loops-, no edge between i and j yet exists -so as to avoid the appearance of multiedges-, and the number of edges which have already been attached to vertex i (j) is smaller than its preassigned degree, k_i (k_j). This process is repeated until the $(\sum_i k_i)/2$ edges are introduced in the network[2]. The connection process can be implemented in practice as follows: The first vertex, i, is selected with a probability proportional to its degree (so that high-degree vertices are more frequently selected), while the second vertex, j, is chosen at random. (This change

[2]When the selection of vertices i and j is carried out at random, it occasionally happens that the execution of the algorithm cannot be completed due to the absence of vertex pairs that satisfy the above three conditions. This occurs usually when the degree distribution is a power law, i. e., when the network contains vertices of large degree. In this case, near the process completion, the vertices which do not have still exhausted their preassigned share of edges are precisely the large-degree vertices, but no new edge can be added to the network because these vertices are usually already connected between themselves. The problem disappears, if we slightly modify the way of selecting the vertex pairs.

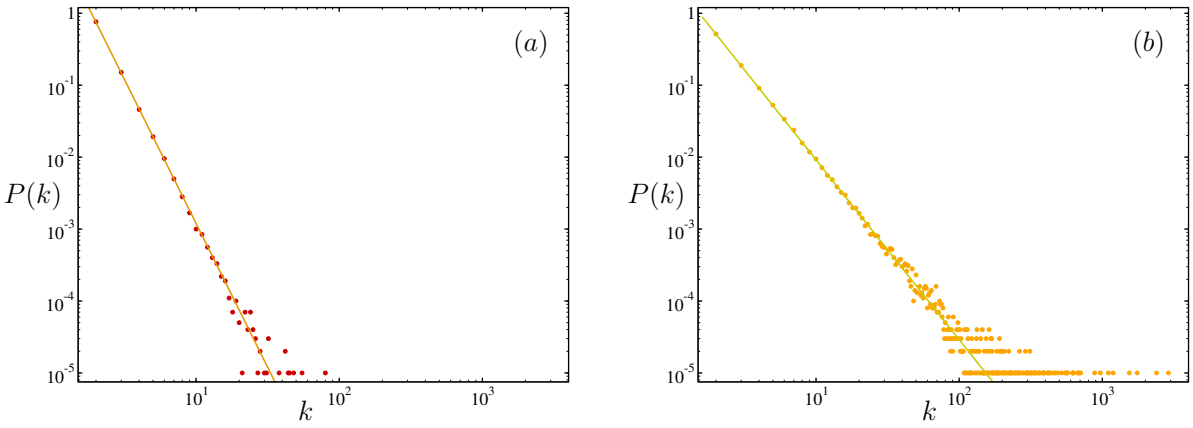

Figure 2.2: Degree distribution $P(k)$ of the *exact scale-free* model *versus* k. Panel (a): $\gamma = 4$; panel (b): $\gamma = 2.5$. The points correspond to the numerical simulations and the curves to equation $P(k) = ak^{-\gamma}$. Both simulations correspond to networks of 10^5 vertices.

certainly generates degree-degree correlations between vertices, but they can be modified later by applying certain algorithms. A discussion about correlations and their effects on networks follows in the next chapter).

To assess the value of the algorithm proposed above, we implement it in a program, and examine the networks thus constructed. In this case, we use it for generating exact scale-free (ESC) networks. Figure 2.2 shows the degree distribution, $P(k)$, corresponding to two ESC networks, one with $P(k) \propto k^{-4}$ (panel a) and another with $P(k) \propto k^{-2.5}$ (panel b). Both networks contain 10^5 vertices. The points correspond to numerical simulations, while the straight lines indicate the slopes -4 (panel a) and -2.5 (panel b). We see that the agreement is excellent. Additional numerical checking demonstrates that both simulated ESC networks are simple graphs.

A very important feature of the ESC model is that the first moment of the degree distribution $\langle k \rangle$ is not independent of N. This can be easily seen using the simple *continuous* approach. Let us consider the following approximation

$$\langle k \rangle \; = \sum_{m}^{N} kP(k) \; \simeq \; \int_{m}^{N} kP(k)dk \; . \tag{2.10}$$

If we now introduce in this equation our scale-free degree distribution $P(k) = ak^{-\gamma}$, we obtain

$$\langle k \rangle \; \simeq \; a \int_{m}^{N} k^{-\gamma+1} dk \; = \; \frac{1}{\gamma - 2} \left[\frac{1}{m^{\gamma-2}} - \frac{1}{N^{\gamma-2}} \right] \; . \tag{2.11}$$

For the case $\gamma = 2.5$ and $m = 2$, the first moment gives $\langle k \rangle \simeq \sqrt{2} - 2/\sqrt{N}$. Thus, small networks of 10^3 vertices will have $\langle k \rangle \simeq 1.351$, while large networks ($N > 10^6$) will have $\langle k \rangle \simeq \sqrt{2} \simeq 1.414$. Hence, as N grows the average number of edges in the network can significantly increase. The effect has a great influence on the topology of the networks; networks generated using the same algorithm but of different orders exhibit non-equivalent topological properties. Note that the referred feature is even more pronounced if one introduces cutoffs in the model, i. e., if the restriction $m \leq k < K$ ($K < N$) is imposed, as done by a number of authors [Catanzaro et al., 2005, Cohen et al., 2000a].

Chapter 3

Vertex-pair correlations in scale-free networks

The degree distribution is one of the principal measures used to capture the topological properties of networks. However, the degree distribution does not suffice to characterize entirely the network topology. Delving a little deeper into the statistics of network structure, one can ask about which vertices pair up with which others. The study of real networks has made clear that the vertex-pair correlations play a key role in the processes taking place in networks. For example, how information or infections spread through a network or how a network collapses in dynamic stages following a failure are processes which cannot completely be understood without taking into account the particular connection probabilities between the vertices. Thus, real networks show vertex-pair correlations, and they are essential to capture the structural features which are responsible (through the connection between structure and behavior) for the dynamic processes which happen in networks. To analyze the role of correlations in networks, we introduce in this chapter three algorithms capable of changing the connection probabilities between vertices in three different ways. They work without varying the degree sequence of the networks considered, so that the structural changes are only due to the vertex-pair correlations. The first two algorithms can change the degree-degree correlations of a given network to respectively produce assortative and dissortative mixing. The third algorithm is able to modifying the transitivity of networks in such a way that one can obtain networks ranging from fully random to maximally triangulated. We also discuss the topological properties that emerge when applying our algorithms to scale-free networks. We state that different vertex-pair correlated networks, in spite of having exactly the same degree sequence, exhibit very different global and transport properties.

3.1 Properties of uncorrelated networks

We begin by revising what we know about uncorrelated networks. A network is said to be degree-degree uncorrelated if the probability that an edge connects to a vertex of a certain degree k is independent from whatever vertex is attached to the other end of the edge. Mathematically, correlations are determined by means of the probability function $\mathcal{E}_{ij}$, which gives the probability that a randomly selected edge of a network connects two vertices, one of degree i and another of degree j. Taking into account that in any network (with a given degree distribution $P(k)$), each end of an edge connects to a vertex of degree k with a probability proportional to $kP(k)$, a network will be uncorrelated if $\mathcal{E}_{ij} \propto iP(i)jP(j)$. This condition is accepted in corresponding literature as the non-correlation requirement for a given network. Note, however, that this

condition can be fulfilled only if no additional structural constraints exist, which limit the connection probabilities between vertices. Sure enough, the prohibition of existing loops and multiedges, as happens for simple graphs, invalidates the preceding non-correlatedness condition. Simply take into consideration that the constraint that no two vertices may be connected by more than one edge imposes fundamental connection restrictions between high-degree vertices of scale-free networks, restrictions which cause a certain "repulsion" between these vertices [Maslov et al., 2004, Newman, 2003a, Boguñá et al., 2004]. This constraint makes impossible that the above condition of uncorrelatedness can be satisfied in many simple graphs. Thus, more precisely, a *pseudograph* is said to be degree-degree uncorrelated if the joint probability that an edge connects two vertices of degrees i and j is:

$$\mathcal{E}_{ij} = \mathcal{E}^r_{ij} := (2 - \delta_{ij}) \frac{iP(i)}{\langle i \rangle} \frac{jP(j)}{\langle j \rangle} \ , \tag{3.1}$$

where $\langle i \rangle = \langle j \rangle$ denotes the first moment of the degree distribution (assumed to be finite). The factor $(2 - \delta_{ij})/\langle i \rangle \langle j \rangle$ emerges because of the normalization requirement[1] ($\sum_{i,j} \mathcal{E}_{ij} = 1$, for $\{i, j : i \leq j\}$ must be satisfied). When a network exhibits different connection probabilities $\mathcal{E}_{ij}$ than those given by Eq. 3.1, then the network is said to be degree-degree correlated. Uncorrelated networks are also called random networks with a given degree distribution, or simply random networks.

While most real networks indeed show the presence of correlations, that is $\mathcal{E}_{ij} \neq \mathcal{E}^r_{ij}$, uncorrelated networks are nevertheless equally important as null models in which dynamical and topological results of correlated networks are compared [Itzkovitz et al., 2003, Milo et al., 2002]; moreover, many analytic solutions are only available in the absence of correlations [Zanette, 2002, May and Lloyd, 2001, Cohen et al., 2000a, Moore and Newman, 2000, Volchenkov et al., 2002, Moreno et al., 2002, Callaway et al., 2000, Pastor-Satorras and Vespignani, 2001, Liu et al., 2003, Olinky and Stone, 2004, Joo and Lebowitz, 2004]. Therefore, it becomes an interesting issue to possibly construct uncorrelated networks of any desired degree sequence (to suitably compare it with the arbitrarily correlated ones), and to satisfy the constraint of no multiedges and loops (because most real networks are simple graphs). As we will soon see, there are different prescriptions for the construction of networks satisfying approximately these properties [Milo et al., 2002, Maslov and Sneppen, 2002, Catanzaro et al., 2005], however, there is no perfect algorithm to generate exactly such random networks [Maslov et al., 2004, Catanzaro et al., 2005, King, 2004, Itzkovitz et al., 2004, Milo et al., 2004, Boguñá et al., 2004].

The classical algorithm used to construct uncorrelated graphs is the so-called configuration model [Bender and Canfield, 1978, Molloy and Reed, 1995, 1998]. Following it, one starts assigning to each one of the N vertices of the system a random number k_i, its degree, drawn from the degree sequence of interest. The assignment of degrees can be represented as an attachment of k_i "edge stubs" (ends of edges incising in the vertex) to each vertex i; considering this, the next step consists of connecting at random pairs of these stubs in order to make complete edges. The only constraint that must be imposed to the process is that the sum $\sum_i k_i$ must, obviously, be even. The result is an uncorrelated network whose degrees are distributed according to

[1] The δ_{ij} appears to avoid the double counting of edges connecting vertices of the same degree. Many authors introduce another quantity (usually denoted $P(i, j)$, see for instance Boguñá and Pastor-Satorras [2003]) to characterize correlations in networks. Using this quantity one must multiply it by $(2-\delta_{ij})$ to obtain the probability that a randomly chosen edge connects two vertices of degrees i and j, that is, to obtain the connection probabilities $\mathcal{E}_{ij}$. This occurs because implicitly in the definition of $P(i, j)$, the edges whose end-vertices have the same degree are counted twice. By the definition of $\mathcal{E}_{ij}$, however, each edge is counted exactly once. The advantage of using $P(i, j)$ is that many expressions related to network correlations can be expressed in a more compact way. Their disadvantage is that it does not represent any "physical" probability, unlike what happens with function $\mathcal{E}_{ij}$. We discard this approach here.

the initially prescribed degree sequence and in which there are no degree correlations, given the random nature of connections. The connection probabilities that the configuration model generates are actually those given by Eq. 3.1. Unfortunately, we observe that the algorithm allows the formation of multiedges and loops; moreover, it generates a non-negligible fraction of them when the degree sequence contains a significant amount of vertices with large degree (hubs). Thus, the configuration model is an algorithm which is perfectly suitable for the construction of random pseudographs with a given degree sequence, but not random simple graphs. We could introduce a third step: if the resulting graph is not a simple graph, it is rejected; otherwise, it is accepted. In this case, the algorithm does correctly generate our searched uncorrelated networks, but the acceptance rate might be too small for the algorithm to be practical [King, 2004].

To obviate this problem Milo et al. [2002] and Catanzaro et al. [2005] suggested the following modification in the algorithm: Instead of pairing up all stubs at once, and then rejecting the graph if it is not simple, they take an incremental approach in which one stub pair is chosen at a time; if the addition of the edge between these two stubs would create a loop or multiedge, the choice is discarded and a new pair is chosen; otherwise, the edge is added to the graph. If at some stage there is no stub pair that can be added without creating a multiedge or loop, the partial graph is rejected and the process is started from scratch. Unfortunately, as King [2004] demonstrates, this modified algorithm does not uniformly generate simple graphs. (Although, it does not result in noticeable biases for large graphs typical of practical studies [Itzkovitz et al., 2004]).

Catanzaro et al. [2005] proposed in their work another modification related to the impossibility (in scale-free networks) of hubs "uncorrelately" connecting other vertices. The effect arises in finite simple networks having heavy-tailed degree sequences due to the constraint that no pair of vertices in a simple graph may be connected by more than one edge; as a consequence of this, hubs must predominantly connect vertices of relative small degree, which can be seen as an "effective repulsion" between hubs [Maslov et al., 2004]. The proposal only differs from the configuration model in the assignment of the degrees: Thus, the degree of each vertex is assigned probabilistically from a desired degree distribution (not from a given degree sequence), which, additionally, is restricted by the constraint $m \leq k_i < N^{1/2}$ [Boguñá et al., 2004] (m and $N^{1/2}$ are, respectively, the minimum and maximum allowed degrees, being N the order of the network). The modification basically eliminates the hubs of the network by introducing the cut-off $N^{1/2}$. However, the elimination of hubs reduces the applicability of the algorithm, which is mainly used to compare different correlated networks with uncorrelated ones, having exactly the same degree sequence. Moreover, hubs play an important role in the topology of real networks, both in a dynamical and structural way; even for simulating theoretical processes taking place in uncorrelated networks hubs should be taken into account.

An alternative approach to these matching algorithms are the switching algorithms, which are based on Markov chains. The idea of these methods is to rewire the edges of a given network until the new resulting network reaches a stationary state in which it shows the desired topological properties [Farkas et al., 2004]. In our case, the algorithm must restructure the connections between vertices until the network becomes uncorrelated. According to this algorithm [Maslov and Sneppen, 2002, Milo et al., 2004, Manna and Kabakçioğlu, 2003], two edges connecting four different vertices are randomly chosen at each step[2] (see fig. 3.1). Then, one end of each edge is selected randomly and the attaching vertices are interchanged. In case one or both of these new edges already exists in the network, this exchange is discarded and a new pair of edges is

[2]In the improbable case, in which the two selected edges have a same end-vertex in common, this pick is discarded and a new pair of edges is chosen; note that in this case the rewiring process could produce a loop.

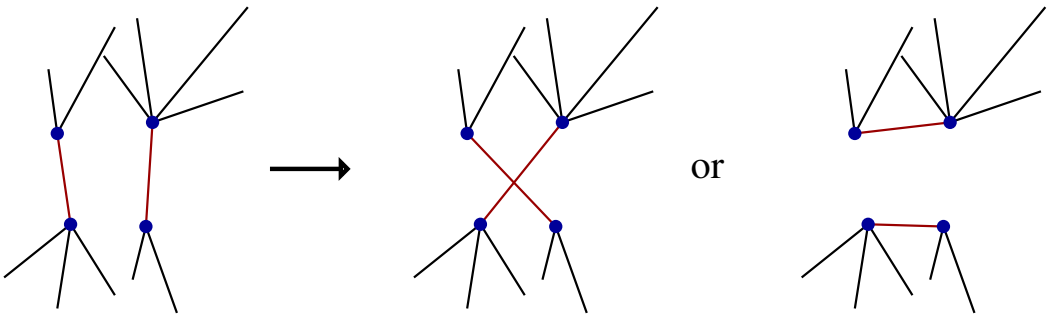

Figure 3.1: Sketch of the edge-restructuring process in the random switching algorithm. Note that, in general, there are two possibilities for rewiring the chosen pair of edges. If the rewiring process links two vertices which are previously already connected, the step is discarded and a new pair of edges is selected.

selected (this restriction prevents the appearance of multiedges). A repeated application of the entire switching process leads to an uncorrelated version of the original network which preserves their structural character of simple graph. Note that the algorithm does not change the degree of vertices involved, and therefore, it preserves the initial degree sequence of the given network. This algorithm works well, but, as with many Markov chain methods, suffers because in general we have no measure of how long we need to wait for it to mix properly. Theoretical bounds on the mixing time exist only for specific near-regular degree sequences [Kannan et al., 1999], but not for scale-free ones. We empirically found that for many scale-free networks a number of iterations approximately equal to 1000 times the number of edges is enough to randomize the network (Milo et al. [2004] find that 100 times is sufficient).

Let us demonstrate analytically that this random switching procedure changes any initial probabilities $\mathcal{E}_{ij}$ of an arbitrarily correlated network into uncorrelated ones: $\mathcal{E}_{ij} = \mathcal{E}^r_{ij}$. To this end we define the function E_{ij}, which gives the number of edges in the network connecting vertices of degree i to vertices of degree j. This new function is related to $\mathcal{E}_{ij}$ as the degree sequence of a network is related to the degree distribution; we empirically determine $\mathcal{E}_{ij}$ averaging the measured E_{ij} over many realizations of the system: then, $\mathcal{E}_{ij} = \langle E_{ij} \rangle / L$, where L is the total number of links in the network. The function E_{ij} has the following immediate properties: 1) $E_{ij} = E_{ji}$, 2) $\sum_{ij} E_{ij} = L$, $\forall ij : i \leq j$, and 3) $\sum_j E_{ij} = iN_i - E_{ii}$, where N_i is the number of vertices of degree i. We proceed next to investigate how the function E_{ij} changes every time the switching procedure is applied. After analyzing all possibilities, one can demonstrate that the function E_{ij} can only increase or decrease by unity, by two, or does not change. As a next step, we calculate the probabilities of change $E_{ij} \to E_{ij} \pm 1$ and $E_{ij} \to E_{ij} \pm 2$. For simplicity, in this calculus we omit the restriction we have imposed in the algorithm to prevent the appearance of multiedges and loops (multiedges and loops are in any case rare in the limit of large networks). Making use of this approximation and of the above properties of E_{ij}, we obtain the following expressions for the probabilities of change:

$$\mathcal{P}^r_{+1} = \begin{cases} \dfrac{(iN_i - 2E_{ii})(iN_i - 2E_{ii} - 1)}{2L(L-1)} & \text{if} \quad i = j \\ \dfrac{(iN_i - E_{ij})(jN_j - E_{ij}) - 4E_{ii}E_{jj}}{L(L-1)} & \text{if} \quad i \neq j \end{cases} \quad , \qquad E_{ij} \to E_{ij} + 1 \ , \tag{3.2}$$

where we denote by $\mathcal{P}^r_{+1}$ the transition probability $E_{ij} \to E_{ij} + 1$, in which E_{ij} increases by

unity,

$$\mathcal{P}^r_{-1} = \begin{cases} \dfrac{2E_{ii}(L - iN_i + E_{ii})}{L(L-1)} & \text{if} \quad i = j \\ \dfrac{E_{ij}(2L - iN_i - jN_j) - 4E_{ii}E_{jj}}{L(L-1)} & \text{if} \quad i \neq j \end{cases} \quad , \qquad E_{ij} \to E_{ij} - 1 \ , \tag{3.3}$$

corresponding to the transition probability $E_{ij} \to E_{ij} - 1$, in which E_{ij} decreases by unity ($\mathcal{P}^r_{-1}$),

$$\mathcal{P}^r_{+2} = \begin{cases} 0 & \text{if} \quad i = j \\ \dfrac{2E_{ii}E_{jj}}{L(L-1)} & \text{if} \quad i \neq j \end{cases} \quad , \qquad E_{ij} \to E_{ij} + 2 \ , \tag{3.4}$$

which is the transition probability ($\mathcal{P}^r_{+2}$) for the case $E_{ij} \to E_{ij} + 2$, and

$$\mathcal{P}^r_{-2} = \begin{cases} 0 & \text{if} \quad i = j \\ \dfrac{E_{ij}(E_{ij} - 1)}{2L(L-1)} & \text{if} \quad i \neq j \end{cases} \quad , \qquad E_{ij} \to E_{ij} - 2 \ , \tag{3.5}$$

which corresponds to the transition probability in which E_{ij} decreases by two unities $E_{ij} \to E_{ij} - 2$. The process of repeatedly applying the switching algorithm corresponds to an ergodic Markov chain, which ensures the existence of a stationary solution. To obtain the solution, we use the transition probabilities above calculated to pose the following two master equations governing the process:

$$\langle E_{ii}(n+1)\rangle = \langle E_{ii}(n)\rangle + \frac{iN_i(iN_i - 1)}{2L(L-1)} + \frac{1 - 2L}{L(L-1)}\langle E_{ii}(n)\rangle \qquad \text{for} \qquad i = j \ , \tag{3.6}$$

and

$$\langle E_{ij}(n+1)\rangle = \langle E_{ij}(n)\rangle + \frac{iN_ijN_j)}{L(L-1)} + \frac{1 - 2L}{L(L-1)}\langle E_{ij}(n)\rangle \qquad \text{for} \qquad i \neq j \ , \tag{3.7}$$

one for the case $i = j$ and another for the case $i \neq j$. These equations say how the expected values of functions E_{ii} and E_{ij}, respectively, are related to two successive iterations n and $n+1$ of the switching process. Note that the equations only depend on L and on the degree sequence of the network, quantities which we know. Dividing both equations by L we obtain

$$\mathcal{E}_{ii}(n+1) = \mathcal{E}_{ii}(n)\left(1 - \frac{2L-1}{L(L-1)}\right) + \frac{iN_i(iN_i - 1)}{2L^2(L-1)} \qquad \text{for} \qquad i = j \ , \tag{3.8}$$

and

$$\mathcal{E}_{ij}(n+1) = \mathcal{E}_{ij}(n)\left(1 - \frac{2L-1}{L(L-1)}\right) + \frac{iN_ijN_j}{L^2(L-1)} \qquad \text{for} \qquad i \neq j \ . \tag{3.9}$$

The last equations have the form $y(n+1) = y(n)(1-a) + b$, whose solution is $y(n) = y_0(1-a)^n + (b/a)[1 - (1-a)^n]$, where y_0 corresponds to the initial value $y_0 = y(n=0)$. Now, using the identities $\sum_i N_i = N$, $P(i) = N_i/N$ and $\langle i\rangle = 2L/N$, and taking into account that in our case $|1 - a| = |1 - (2L-1)/(L(L-1))| < 1$, we can calculate the expected correlations. In the limit of large networks $N \to \infty$, $L \to \infty$ (in which we are interested) and assuming an infinite number of iterations $n \to \infty$, we have:

$$\mathcal{E}_{ij} = \begin{cases} \left(\dfrac{iP(i)}{\langle i\rangle}\right)^2 & \text{if} \quad i = j \\ \dfrac{2iP(i)\ jP(j)}{\langle i\rangle\langle j\rangle} & \text{if} \quad i \neq j \end{cases} \qquad \longrightarrow \qquad \mathcal{E}_{ij} = (2 - \delta_{ij})\frac{iP(i)jP(j)}{\langle i\rangle\langle j\rangle} \ . \tag{3.10}$$

This result is exactly the same as Eq. 3.1, which proves that this switching procedure would generate exactly an uncorrelated pseudograph if the restrictions imposed, in oder to avoid the appearance of multiedges and loops, are excluded. Note also that the solution does not depend on y_o, which confirms that the final probabilities $\mathcal{E}_{ij}$ produced by the algorithm are independent of the initial correlations of the network. On the other hand, one can observe, after analyzing the equations (3.8) and (3.9), that the system must rapidly converge to the stationary state.

Eq. 3.10 proves that the random switching algorithm (RS) works very successfully for producing uncorrelatedness. Of course, when the restrictions to avoid the appearance of loops and multiedges are imposed, the resulting connection probabilities will be slightly different from those given by Eq. 3.1, most notably when the degrees i and j are large, as Boguñá et al. [2004] points out. This does not mean that the algorithm is unsuccessful for constructing uncorrelated scale-free simple graphs. The RS algorithm does not achieve the production of such graphs because it is structurally impossible that hubs connect "uncorrelately" between themselves (to the effect of Eq. 3.1) in scale-free networks [Maslov et al., 2004, Boguñá et al., 2004, Catanzaro et al., 2005]. In our opinion, the problem is actually a question of definition: Certainly, an operative definition of what an uncorrelated pseudograph is (Eq. 3.1) exists, but there is no definition to explain what an uncorrelated simple graph is. In this respect, we argue that the RS algorithm generates what could properly be defined as uncorrelated simple graphs, given that the generated connections between vertices do not depend on degrees or other properties of the involved vertices (as a result of the random nature of the rewiring process), and that they satisfy the structural restrictions which are characteristic of simple graphs (no loops and multiedges).

Let us now introduce some quantities to measure the uncorrelatedness. Several different ways of quantifying degree-degree correlations have been proposed in corresponding literature. One possibility is simply to plot the two-dimensional function E_{ij} (or $\mathcal{E}_{ij}$), that is, the histogram of vertex degrees at either ends of an edge [Maslov and Sneppen, 2002, Maslov et al., 2004]. Unfortunately, this direct measurement is usually a rather complex task owing to large statistical fluctuations. A more compact representation to evaluate correlations consists of calculating the so-called nearest neighbors' average degree function [Pastor-Satorras et al., 2001, Vázquez et al., 2002, Xulvi-Brunet and Sokolov, 2005], which, in terms of the introduced connection probabilities $\mathfrak{E}_{ij}$, is defined as

$$\bar{k}_{nn}(i) = \frac{\sum_j j(1+\delta_{ij})\mathcal{E}_{ij}}{\sum_j (1+\delta_{ij})\mathcal{E}_{ij}} \ . \tag{3.11}$$

This function gives a one-parameter curve which, for uncorrelated pseudographs, is equal to the constant $\bar{k}^r_{nn}(i) = \langle i^2\rangle/\langle i\rangle$, independent of the degree i. (Note that it is made necessary to introduce the factors $1+\delta_{ij}$ in Eq. 3.11 if we want to have $\bar{k}_{nn} =$ constant for the uncorrelated case. The measurement of correlations can still be further reduced to single numbers by calculating either the Pearson correlation coefficient of the degrees at the endpoints of an edge [Newman, 2002, 2003a], or the following parameter [Newman, 2003a, Xulvi-Brunet and Sokolov, 2004]:

$$\mathcal{A} = \frac{\sum_i \mathcal{E}_{ii} - \sum_i \mathcal{E}^r_{ii}}{1 - \sum_i \mathcal{E}^r_{ii}} \ . \tag{3.12}$$

This simple coefficient vanishes when the network is uncorrelated, while it takes values different from zero when the network presents correlations. The one-parameter $\bar{k}^r_{nn}(i)$, Eq. 3.11, naturally reveals much more information about the connection probabilities than the Pearson coefficient or parameter $\mathcal{A}$. (Note that two different connection probability functions, $\mathcal{E}^1_{ij}$ and $\mathcal{E}^2_{ij}$, could provide the same value for $\mathcal{A}$ or the Pearson parameter.)

Let us examine the degree-degree correlations of RS networks with the aid of the above described quantities. To this end, we consider Barabási-Albert networks with $L = 2\cdot N$ ($m = 2$)

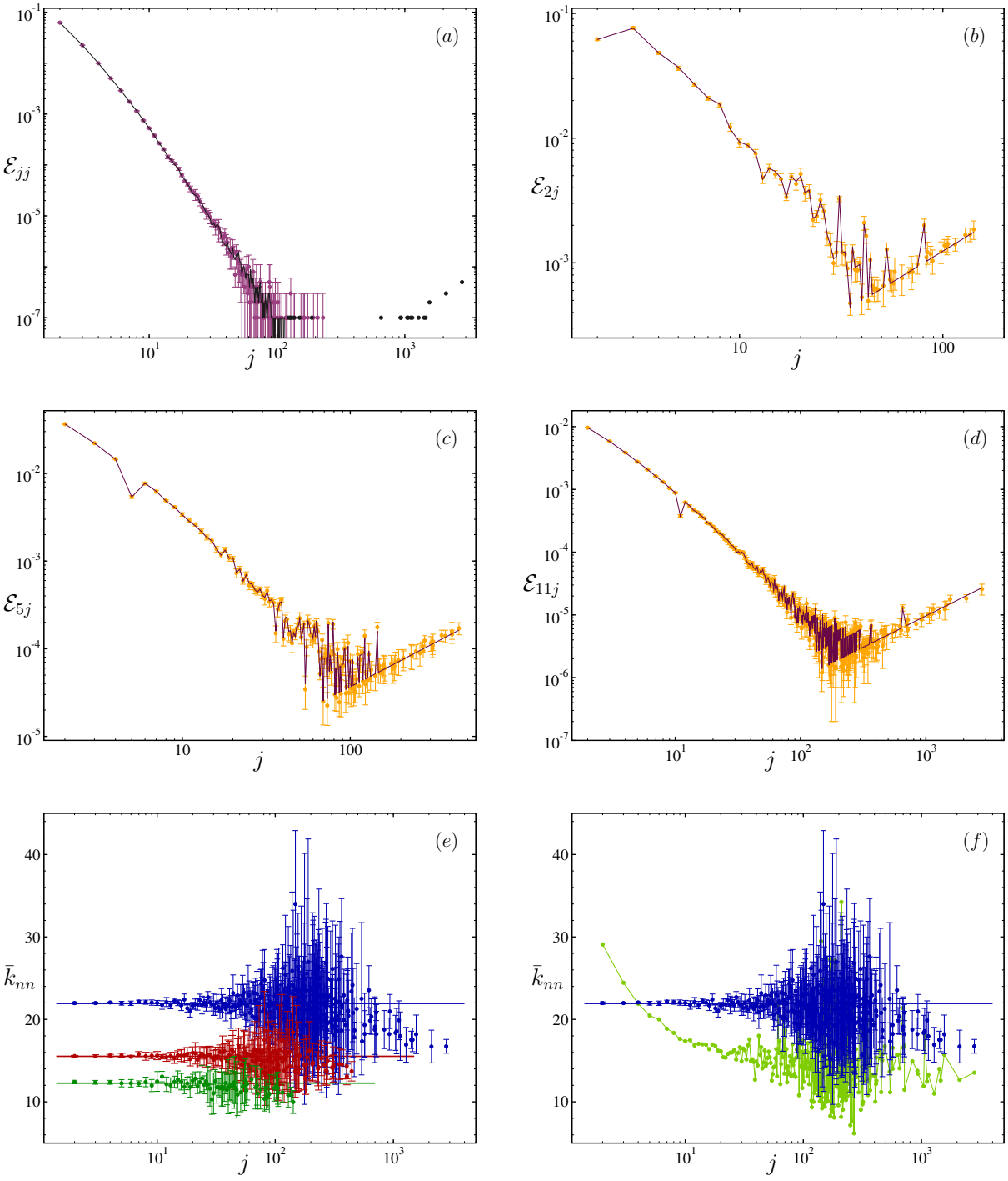

Figure 3.2: (a)-(d): Different connection probabilities, $\mathcal{E}_{ij}$, of several scale-free networks (see text for more details about the networks) to which the *RS* algorithm has been applied versus the degree j. The points with error bars correspond to the measured $\mathcal{E}_{ij}$. The curves correspond to the theoretical $\mathcal{E}^r_{ij}$, Eq. 3.1; they have been calculated using $P(k) = N_k/N$, where N_k is the degree sequence of the network considered. (e): nearest neighbors' average degree function, $\bar{k}_{nn}$, for different uncorrelated networks generated using the *RS* algorithm. (f): Comparison between the nearest neighbors' average degree functions of an original Barabási-Albert network (bottom) and the same network after applying the *RS* algorithm (top).

as initial networks to which the algorithm is applied. The algorithm is applied ten times to each network; all results we present are averaged over the ten realizations. We begin investigating the connection probabilities. Figure 3.2 shows $\mathcal{E}_{ij}$ for different values of indices i and j of these networks: panel (a) shows $\mathcal{E}_{jj}$ as a function of the degree j for a network of 10^6 vertices; panel (b) shows $\mathcal{E}_{2j}$ as function of the degree j for a network of 10^4 vertices; panel (c) shows $\mathcal{E}_{5j}$ as function of the degree j for a network of $3 \cdot 10^4$ vertices; and panel (d) shows $\mathcal{E}_{11j}$ as function of the degree j for a network of again 10^6 vertices. The points represent the measured $\mathcal{E}_{ij}$ in the four graphs. The plotted curves correspond to the theoretical connection probabilities, $\mathcal{E}^r_{ij}$, given by Eq. 3.1; the curves have been calculated using the particular degree distribution $P(k) = N_k/N$ of the network considered. We note that agreement is excellent. This is not surprising, since the probability that multiedges and loops appear in a large network is almost in all cases negligible. One case in which this probability may not be neglected is precisely when the number of vertices of degree k is much smaller than k, as occurs for hubs in scale-free networks. In effect, an accurate analysis of the connection probabilities between hubs indicates a small disagreement with Eq. 3.1 (shown in panels e and f). This already expected behavior, however, has a negligible influence on the whole distribution of connections, since the fraction of edges which are attached to hubs is actually very small. Thus, the pictures make clear that the RS algorithm is certainly able to change the initial $\mathcal{E}^{in}_{ij}$ into uncorrelated ones.

Panel (e) of figure 3.2 shows the nearest neighbors' average degree function, $\bar{k}^r_{nn}(j)$, versus the degree j for the above networks. The results correspond, from bottom to top, to the network having 10^4, $3 \cdot 10^4$, and 10^6 vertices. The points (with the corresponding error bars) represent the measured $\bar{k}^r_{nn}(j)$, while the straight lines point at the corresponding constants $\langle j^2 \rangle / \langle j \rangle$. The picture demonstrates that the engendered connection probabilities coincide exactly with those given by Eq. 3.1 except for the ones corresponding to edges attached to very large-degree vertices. The bias gives evidence that hubs connect "reluctantly" to other hubs, according to the "effective repulsion" produced by the constraint that no two vertices are connected by more than one edge. We want, however, again to point out that hubs represent a very small fraction of the totality of vertices. For example, the network of $N = 10^6$ has only 61 vertices of degree larger than $j = 300$, which are, roughly, the vertices that seem, in the picture, to connect "in a correlated way". However, this number entails only the extremely small fraction of 6^{-5} vertices. In panel (f) we compare the nearest neighbors' average degree functions, $\bar{k}_{nn}$, corresponding to: (green) the original Barabási-Albert network of $N = 10^6$ (which is slightly correlated), and the same network after applying the RS algorithm. The notable difference between the drawn curves shows the sensibility of the one-parameter function $\bar{k}_{nn}$ for detecting correlations.

The preceding simulation results demonstrate the reliability of the RS algorithm. The global connection probabilities, $\mathcal{E}_{ij}$, of the RS networks do not noticeably differ from the uncorrelated ones, Eq. 3.1, even in the case in which the networks have heavy-tailed degree sequences. Thus, aside from theoretical discussions about the suitability of the definition of "uncorrelatedness" on simple graphs, the present analysis points out the validity of the RS algorithm for the construction of uncorrelated networks (to the effects of Eq. 3.1) in a practical way. We will use it to build our random networks with given degree sequences along this thesis.

3.2 Assortativity and dissortativity

The random-correlated model has played in the last few years, and still plays, an important role in the study of networks. However, real-world networks do exhibit correlated mixing between vertices. They show a complex mixing of vertex-pair connections so that the joint probability function $\mathcal{E}_{ij}$ satisfies $\mathcal{E}_{ij} \neq \mathcal{E}^r_{ij}$. For instance, in many social networks, like the different

co-authorship networks or the film actor collaborations network, high-degree vertices attach preferably to other highly connected vertices [Newman, 2003a, 2002]. In current literature, this characteristic is referred to as *assortativity* or *assortative* mixing. On the other hand, technological and biological networks show the property that vertices having high degrees tend to connect to low-degree vertices, and vice versa. This characteristic is usually called *dissortativity*. Examples of networks exhibiting *dissortative* mixing are the Internet, the World Wide Web, power grids, protein-protein interactions, or the metabolic network [Maslov and Sneppen, 2002, Newman, 2003a]. Moreover, most theoretical models dealing with evolving networks show that correlations appear spontaneously due to the organizing principles which guide their growth [Krapivsky and Redner, 2001, Catanzaro et al., 2004]. It seems then that correlations are an intrinsic characteristic of networks.

In order to assess the exact influence of correlations on network topology, several authors have proposed procedures to build correlated networks. In a recent study Ramezanpour and Karimipour [2003] show that the transformation which converts a random graph with a specific degree distribution to their line (or edge-dual) graph produces a new graph with nearly the same degree distribution, but with degree correlations and higher clustering coefficient. The line graph is constructed as follows: to each edge of the initial network, a vertex of the line graph is assigned; any two vertices of the line graph are connected by an edge if the corresponding edges of the original network are incident on the same vertex. On the other hand, Xulvi-Brunet et al. [2003] demonstrate that a minor change in the RS algorithm leads to a network which is assortative: At each step, instead of choosing two edges randomly, two vertices are selected at random, and then one incident edge from each vertex is picked randomly; finally, the two edges chosen in such a way are rewired following the same procedure as described in the RS algorithm. Both preceding models, like others [Catanzaro et al., 2005], suffer from a certain lack of flexibility since they are not able to construct networks with both the wanted degree sequence and the desired degree of "correlatedness".

The most general procedures have been proposed by Boguñá and Pastor-Satorras [2003] and Newman [2003a], who suggested two different ways to construct general correlated networks with given correlations. Boguñá and Pastor-Satorras [2003] propose the following algorithm: firstly, assign to each vertex i a degree k_i drawn from the wanted degree distribution $P(k)$, and secondly, for each pair of vertices i and j, draw an edge between them with the suitable probability for generating the desired $\mathcal{E}_{ij}$. However, the algorithm is not easy to implement numerically because of the small probabilities involved when it is applied to scale-free networks. Moreover, for aiming to compare correlated networks the model is not perfect, given that the desired degree (N_k) and the degree-degree (E_{ij}) sequences cannot be exactly generated. On the other hand, Newman proposes to start from a given degree-degree sequence E_{ij} from which one calculates the corresponding degree sequence N_i of the network, in order to afterwards generate an uncorrelated simple graph using the modified configuration model or the switching algorithm. The next step consists of rewiring of edges: one measures the degrees of the vertices associated with the two randomly selected edges, and then rewires them or not if a given condition is satisfied [Newman, 2003a].

To construct networks using the last two algorithms, it is necessary to choose in advance the degree-degree sequence function E_{ij} (or the connection probabilities $\mathcal{E}_{ij}$). The selection of E_{ij} is, however, not easy, due to the fact that we do not usually know which function E_{ij} exactly corresponds to what properties showed by networks. Therefore, we propose a different perspective for approaching the question [Xulvi-Brunet and Sokolov, 2004]. Instead of putting in correlations "by hand", we suggest to proceed by doing the opposite: Imposing the desired conditions which must be satisfied, that is, that "vertices with similar degree must preferably connect between them" (assortativity) and "low-degree vertices try to connect with high-degree

vertices" (dissortativity), and then analyzing the corresponding correlations which come out of the models. For this purpose, we change the connections between vertices by means of new rewiring processes which also preserve the degree sequence of the networks. Thus, we present here two algorithms working under the above guidelines to construct, respectively, assortative and dissortative networks having the degree of assortativity (dissortativity) we desire. The idea behind this study is to check what the effects of assortativity (dissortativity) alone on the properties of networks are, which stay random in any other respect. Thus, the models we propose here do not take into account some of the important effects which, for instance, arise from geographical restrictions or simply peculiar properties of certain vertices in the network. However, we will see that these simple mathematical models exhibit interesting properties which very probably are also pertinent to real-world networks.

The assortative model. We start from a given network having a desired degree sequence (regardless of which type of correlations it has). The idea is how the initial degree-degree sequence function $E_{ij}^{(in)}$ can be modified so that the final network shows the wanted degree of assortative mixing. In order to do this, we primarily choose at random two edges of the network connecting four different vertices. (As in the RS algorithm, if the two selected edges connect only three vertices, the choice is rejected). Now, we consider the four vertices associated with these two edges, and order them with respect to their degrees. Then, with probability p, we rewire the edges in such a way that one edge connects the two vertices with the smaller degrees and the other connects the two vertices with the larger degrees; otherwise, the edges are randomly rewired. In the case that one or both of these new edges already exist in the network, the step is discarded and a new pair of edges is chosen. The repeated application of the rewiring step leads to an assortative version of the original network. By changing the parameter p, it is possible to obtain networks with different degrees of assortativity, ranging from fully random ($p = 0$) to totally assortative ($p = 1$).

Starting from this assortative switching (AS) procedure, we can obtain a theoretical expression for $\mathcal{E}_{ij}^{a}$ as a function of the parameter p. The restriction $i \leq j$ can be imposed without loss of generality, since undirected networks satisfy $E_{ij} = E_{ji}$ (we shall count each edge of the network only once.) Then, we go on in the same way in which we went to calculate the connection probabilities $\mathcal{E}_{ij}^{r}$ of the RS algorithm: We study how the system varies when the exchange step is carried out, and then write down the master equation which governs this change. In this case, however, we do not concentrate directly on the change of $\mathcal{E}_{ij}$ but on the change of the following variable:

$$F_{ln} = \sum_{r=l}^{n} \sum_{s=r}^{n} E_{rs} \qquad r \leq s \;\; ; \;\; l \leq n \quad . \tag{3.13}$$

A careful analysis of the algorithm reveals that every time the rewiring procedure is applied, this variable F_{ln} either increases or decreases by unity, or does not change. We calculate the probabilities of change $F_{ln} \to F_{ln} + 1$ and $F_{ln} \to F_{ln} - 1$. (In this computation, we consider the possibility that multiedges and loops appear in the system. The reason for this is that this condition simplifies considerably the calculations. Since the probability that loops and multiedges appear in large systems is very small, the solution will not be exact, but a very good approximation for simple graphs.) Now, taking all corresponding possibilities into account, one can obtain the following expressions for the probabilities of change:

$$\mathcal{P}_{+1}^{a} = 2\left[(X_{ln} - f_{ln})\,(X_{ln} - f_{ln} - 1) + p\,(X_{ln} - f_{1n} + f_{1,l-1})^2\right] \qquad F_{ln} \to F_{ln} + 1 \tag{3.14}$$

$$\mathcal{P}_{-1}^{a} = 2f_{ln}\left[(1-p)(1-2X_{ln}) + p(2X_{1,l-1} - f_{1,l-1} - f_{1n}) + f_{ln}\right] \qquad F_{ln} \to F_{ln} - 1 \tag{3.15}$$

Here, $f_{ln} = F_{ln}/L$ (L is the total number of edges in the network), and the function X_{ln} is given by the expression:

$$X_{ln} = \frac{1}{\langle k \rangle} \sum_{k=l}^{n} kP(k) \qquad\qquad l \leq n \ . \tag{3.16}$$

In the last equation we use again $P(k) = N_k/N$, where N_k is the number of vertices of degree k and N the total number of vertices in the network. (The solution could naturally be expressed as a function of the specific degree sequence N_k of the given network. However, because the final formulas are usually expressed as a function of the degree distribution in current literature, we use here $P(k)$). Note that X_{ln} and f_{ln} vanish when one of the indices is smaller than 1, the minimal tolerated degree[3]. Using Eq. 3.14 and Eq. 3.15, we can formulate the equation governing the process (which again corresponds to an ergodic Markov chain), and, in this way, calculate the expected value of the variable f_{ln}, $\forall l, n$. Defining $\langle f_{ln} \rangle = \mathcal{F}_{ln}$, and assuming that $X_{ln} - f_{ln} \gg 1$ (which is always a good approximation), the stationary solution is given by the conditions:

$$\left(X_{ln} - \overset{a}{\mathcal{F}}_{ln}\right)^2 + p\left(X_{ln} - \overset{a}{\mathcal{F}}_{1n} + \overset{a}{\mathcal{F}}_{1,l-1}\right)^2 = \tag{3.17}$$

$$= \overset{a}{\mathcal{F}}_{ln}\left((1-p)(1-2X_{ln}) + p(2X_{1,l-1} - \overset{a}{\mathcal{F}}_{1,l-1} - \overset{a}{\mathcal{F}}_{1n}) + \overset{a}{\mathcal{F}}_{ln}\right) \ ,$$

for all $l > 1$, and

$$(1+p)\left(X_{1n} - \overset{a}{\mathcal{F}}_{1n}\right)^2 = (1-p)\overset{a}{\mathcal{F}}_{1n}\left(1 - 2X_{1n} + \overset{a}{\mathcal{F}}_{1n}\right) \ , \tag{3.18}$$

for $l = 1$. Note that Eq. 3.17 reduces to Eq. 3.18 when $l = 1$. The last equations are sufficient to calculate $\mathcal{F}_{ln}$, for all possible values of l and n. Resolving them, we obtain, after some calculations, the expression:

$$\overset{a}{\mathcal{F}}_{ln}(p) = \begin{cases} \dfrac{X_{ln}^2 + (B_n - B_{l-1})^2/p}{(1-p)/2 + pX_{ln} + B_n + B_{l-1}} & \text{if} \quad 0 < p \leq 1 \\ X_{ln}^2 & \text{if} \quad p = 0 \end{cases} \ , \tag{3.19}$$

where (by definition, see Eq. 3.13) the restriction $l \leq n$ must be imposed[4]. Here, we denote by B_n the following square root

$$B_n = \sqrt{\left[pX_{1n} + \frac{1-p}{4}\right]^2 - \frac{p(1+p)}{2}X_{1n}^2} \ . \tag{3.20}$$

We can observe that $\mathcal{F}_{ln}$ only depends on the degree sequence of the network (hidden in X_{ln}) and on the algorithm's parameter p, as we have expected. Finally, we calculate the expected connection probabilities $\mathcal{E}_{ij}$. Taking into account Eq. 3.13, we have (as a function of the above $\mathcal{F}_{ln}$):

$$\mathcal{E}_{ij}^{a}(p) = \overset{a}{\mathcal{F}}_{ij}(p) - \overset{a}{\mathcal{F}}_{i,j-1}(p) - \overset{a}{\mathcal{F}}_{i+1,j}(p) + \overset{a}{\mathcal{F}}_{i+1,j-1}(p) \ . \tag{3.21}$$

[3]We do not consider vertices of degree zero, since no edges connect to degree-zero vertices and, therefore, information about isolated vertices is not contained in the edge distribution E_{ij}. The amount of vertices of degree zero must be specified separately.

[4]The solution given by Eq. 3.19 differs slightly from that offered in [Xulvi-Brunet and Sokolov, 2004] and [Xulvi-Brunet and Sokolov, 2005]. The disagreement is due to an unfortunate misprint. In both articles we forgot the p which divides $(B_n - B_{l-1})^2$. The typo does not produces important biases, since the results one can obtain using the formula given in the articles and using Eq. 3.19 are numerically more or less the same. We thank Bartłomiej Wacław for bringing this mistake to our attention.

From Eqs. 3.21 and 3.19, one can prove that this solution reduces to the corresponding uncorrelated case $\mathcal{E}^r_{ij}$ when $p = 0$, and that it takes the form

$$\mathcal{E}^a_{ij} = \delta_{ij} \frac{iP(i)}{\langle i \rangle} \tag{3.22}$$

when $p = 1$. This case corresponds obviously to the totally assortative network, since Eq. 3.22 indicates that all vertices of degree i connect only to vertices of the same degree i, $\forall i$. (On scale-free simple graphs, however, this fully assortative behavior will presumably not happen. Hubs, because of the constraint that no pair of vertices may be connected by more than one edge, will surely connect to vertices of smaller degree. Bear in mind that our solution is strictly valid for pseudographs only, or, at best, for simple graphs in the thermodynamic limit.) We prove then that the algorithm is capable of restructuring the links of any given network in such a way that, as function of the parameter p, the resulting network has the desired degree of assortativity. Additionally, it produces assortative mixing so that connection probabilities between vertices of same degree are still random. This feature, which ensures that the algorithm generates assortative mixing leaving random any other properties of the network, allows us to consider this algorithm as an instrument to construct networks with "random assortative" correlations.

When one speaks of assortativity in networks, one means that high-degree vertices preferably connect other high-degree vertices. But, if vertices having large degrees are preferably connected between themselves, then the rest of the vertices (low- and medium-degree vertices) must also connect preferably to each other. Hence, a more precise definition of assortativity is "vertices with similar degrees tend to connect with a larger probability than in uncorrelated networks", i. e., for assortative networks $E_{ii} > E^r_{ii}$, $\forall i$ [Xulvi-Brunet and Sokolov, 2004]. Of course, it is possible to have a network in which almost all vertices of degree i connect to vertices of degree $i + 1$ or $i - 1$ but not of i, so that $E_{ii} < E^r_{ii}$ is satisfied; following then the previous definition, we will not have an assortative network in this case, in spite of the given intuitive condition for assortativity (similar-degree vertices connect preferably). It could also be that vertices of degree i connect preferably only to vertices having either the same degree i or a very different degree k, such that $k \ll i$ or $k \gg i$; the network could then satisfy the relation $E_{ii} > E^r_{ii}$ (and, therefore, be assortative according to our definition), but, however, it does not satisfy the intuitive idea of assortativity. Thus, we actually expect that real networks are "well-behaved" to the effect that if similar-degree vertices connect preferably, then it corresponds to $E_{ii} > E^r_{ii}$, and vice versa.

The discussion above convinces us of the difficulty to define a good parameter to quantify the degree of assortativity of a network (and the idea of assortativity itself). To measure degree-degree correlations, we can naturally use the different quantities we have already introduced in the last section, but it must be clear that none of them is perfect. For example, take the very simple parameter given by Eq. 3.12; it takes the value 0 when the network is uncorrelated and the value 1 when the network is totally assortative. This parameter (like other ones) is unable to quantify the amount of connections existing between vertices of degree i and vertices having other similar degrees. Thus, if we want to analyze this aspect of our resulting networks, we must introduce another parameter. In general, we shall not focus on one way only to measure assortativity, but we will make use of different quantities throughout the text.

Let us now apply our algorithm, using different values of p, to a scale-free network constructed by following the prescription of Barabási and Albert [Barabási and Albert, 1999]. The network will then present a degree sequence which approximately falls as Eq. 1.6 indicates. Additionally, we take again $m = 2$, corresponding to Barabási-Albert networks which have twice as many numbers of edges than vertices $L = 2N$. In the present study we refrain from discussing the peculiar case $m = 1$, since for $m = 1$ the Barabàsi-Albert construction is a tree, which is destroyed by the application of the algorithm and transformed into a set of disconnected

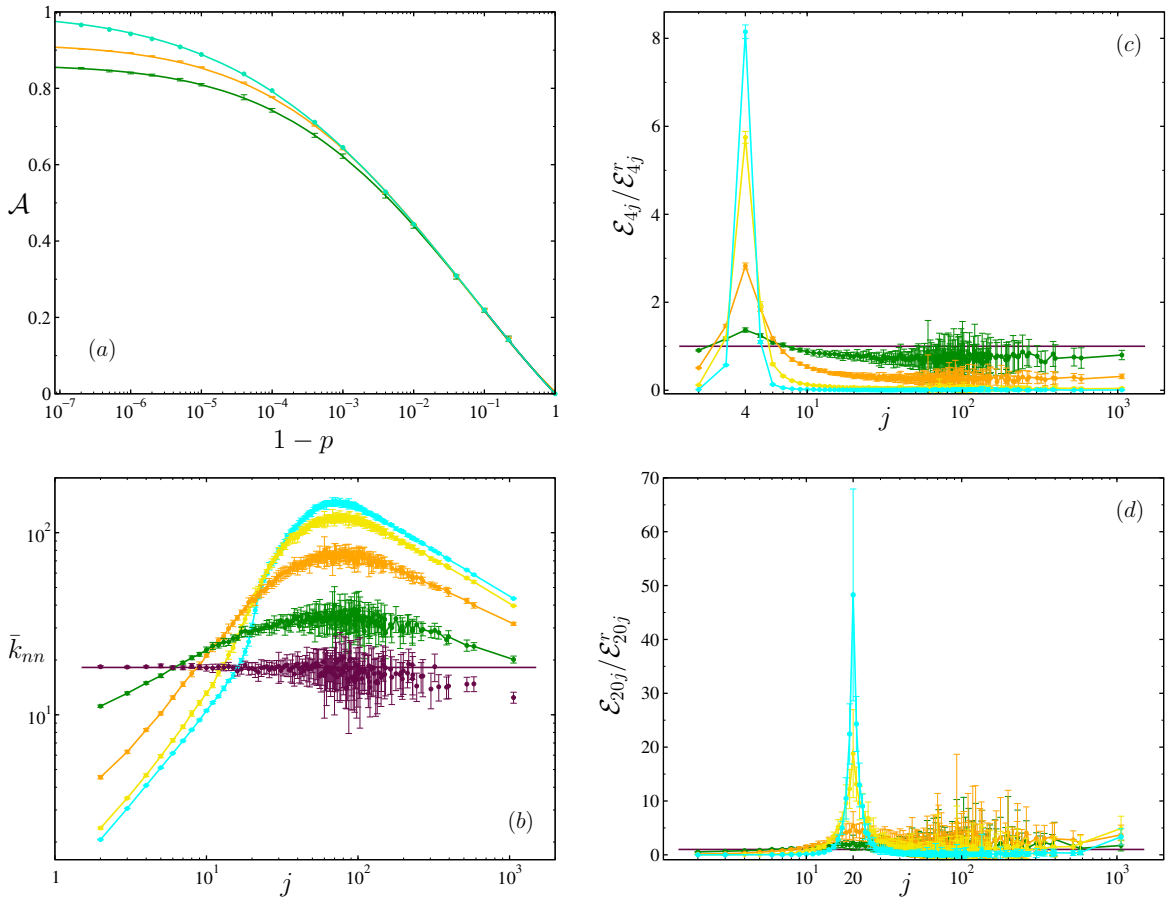

Figure 3.3: Panel (a): The coefficient of assortativity $\mathcal{A}$ as a function of algorithm parameter p (see text). Panel (b): Behavior of function $\bar{k}_{nn}(j)$ versus j for different assortative networks: $\mathcal{A} = 0$ (maroon), $\mathcal{A} = 0.067$ (green), $\mathcal{A} = 0.221$ (orange), $\mathcal{A} = 0.443$ (yellow), and $\mathcal{A} = 0.640$ (cyan). Panel (c): $\mathcal{E}_{4j}/\mathcal{E}^r_{4j}$ as function of j, and panel (d): $\mathcal{E}_{20j}/\mathcal{E}^r_{20j}$ as function of j, both for the same assortative versions studied in panel (b).

clusters. On the other hand, for $m \geq 2$ the same qualitative results were observed. All results we will present in this section are averaged over ten independent realizations of the algorithm as applied to the same original network, so that all assortative versions have exactly the same degree sequence.

In order to measure the assortativity degree of the new versions, we consider primarily the parameter $\mathcal{A}$, Eq. 3.12. To this end, we measure the degree-degree function, E_{ij}, of each assortative version (all versions are generated using the same p), and use the results to calculate the corresponding values of $\mathcal{A}$. The procedure is repeated for all values of p considered. Figure 3.3 (a) shows the variation of the coefficient $\mathcal{A}$ as function of the parameter p. The two lower curves correspond to the measured assortativity of two networks of different order, $N = 10^4$ (low) and $N = 10^5$ (middle), and the upper curve to our theoretical prediction, Eq. 3.21, corresponding to an infinite network with the degree distribution given by Eq. 1.6. We see that all curves coincide for small values of $\mathcal{A}$. However, whereas the theoretical curve reaches the value $\mathcal{A} = 1$ when $p \to 1$, the measured $\mathcal{A}$ corresponding to the finite sized networks increases up to a maximal value which always remains below unity. Sure enough, the central curve of the graph, which

corresponds to the network having 10^5 vertices, reaches the maximal value $\mathcal{A} = 0.917$ when $p \to 1$. For the lower curve, corresponding to the network with $N = 10^4$ vertices, this value is still smaller: $\mathcal{A} = 0.864$. This was already expected and is due to the finite-sized effects mentioned above. The constraint that no pair of vertices is connected by more than one edge bounds $\mathcal{A}$ from above by the values lower than 1. Note, however, that measured $\mathcal{A}$ only differs from the theoretical prediction ($N \to \infty$) when the network is very assortative. For typical degrees of assortativity ($\mathcal{A} \leq 0.6$) the agreement is perfect even if the network is small.

Figure 3.3 (b) represents the function $\bar{k}_{nn}(j)$, the nearest neighbors' average degree, Eq. 3.11, as function of j. The curves correspond to different correlated versions of the same Barabàsi-Albert network (10^5 vertices), each one with a different degree of assortativity. The lower curve corresponds to the randomized version ($\mathcal{A} = 0$), and the straight line to the constant $\langle j^2 \rangle / \langle j \rangle$, calculated with the particular degree sequence of the initial network. The rest of the curves, from bottom to top, corresponds to the assortative versions: $\mathcal{A} = 0.067$, $\mathcal{A} = 0.221$, $\mathcal{A} = 0.443$, and $\mathcal{A} = 0.640$. If a network exhibits assortative mixing we expect that the functions $\bar{k}_{nn}$ are increasing. Sure enough, if vertices connect to vertices of similar degree, then the values of the function $\bar{k}_{nn}$ must be small if k is small (low-degree vertices) and larger as k grows. We observe this increasing behavior for all sufficiently small values of j in all plotted curves. The final decreasing tail of curves appears because of the restriction that multiedges are forbidden in simple networks: Since hubs may not connect to other hubs through more than one edge, most of their edges must attach to vertices of large- or middle-degree (not to low-degree vertices, because in assortative networks low-degree vertices connect between themselves). Note that, the fact that $\bar{k}_{nn}(j)$ satisfies $\bar{k}_{nn}(j) > \bar{k}^r_{nn}(j)$ for all large j, confirms this assumption and, consequently, indicates a certain assortative behavior of hubs (namely, that hubs tend to connect to the vertices which have the larger degree of the network, but regard the restrictions imposed by the simple graph character of the network). Thus, this decreasing tail is actually expected for all scale-free simple networks and is due to finite size effects. (We want again to remark that the fraction of hubs is very small in a scale-free network. Thus, in spite of the apparently large decreasing tails plotted in the picture, the fraction of involved vertices is really negligible).

Panels (c) and (d) of 3.3 show, respectively, the fractions $\mathcal{E}_{4j}/\mathcal{E}^r_{4j}$ and $\mathcal{E}_{20j}/\mathcal{E}^r_{20j}$ for networks of different degree of assortativity, as a function of j. The curves correspond to the same networks we used to study the function $\bar{k}_{nn}$ in panel (b). From bottom to top (taking the peaks of the curves into account): $\mathcal{A} = 0$, $\mathcal{A} = 0.067$, $\mathcal{A} = 0.221$, $\mathcal{A} = 0.443$, and $\mathcal{A} = 0.640$. The curve which represents the behavior of $\mathcal{A} = 0$ is, of course, the straight line corresponding to the constant value equal to one. The behavior changes, as expected, for the assortative networks. We see that they show a peak for $j = 4$ ($j = 20$), and that this peak grows as the assortativity increases. This result is obviously due to the fact that the effect of the algorithm consists of increasing the probability $\mathcal{E}_{ii}$, $\forall i$, as $p \to 1$. More interesting is, however, that the algorithm works in such a way that the probability of a vertex of degree i connecting to a vertex of degree j, decreases gradually as the difference between the degrees i and j ($|i - j|$) grows. To put it another way, vertices of similar degrees certainly connect with larger probability than vertices whose degrees are strongly different. This can easily be observed in the panels: The fraction $\mathcal{E}_{4j}/\mathcal{E}^r_{4j}$ (panel c) is larger than one for values of j close to the selected $j = 4$, which indicates that vertices of degree close to $j = 4$ connect preferably to vertices of degree 4, compared with uncorrelated networks. On the other hand, for j distant from the selected $j = 4$, the considered fraction is smaller than one and it decreases continually as the difference $|j - 4|$ grows. The same behavior can be observed in panel (d). This panel shows also the "simple graph effects" due to the "repulsion" between hubs: Large-degree vertices connect for the most part to large- or middle-degree vertices in assortative networks. Consequently, large- and middle-degree vertices must connect to hubs with larger probability than in uncorrelated networks. This fact explains

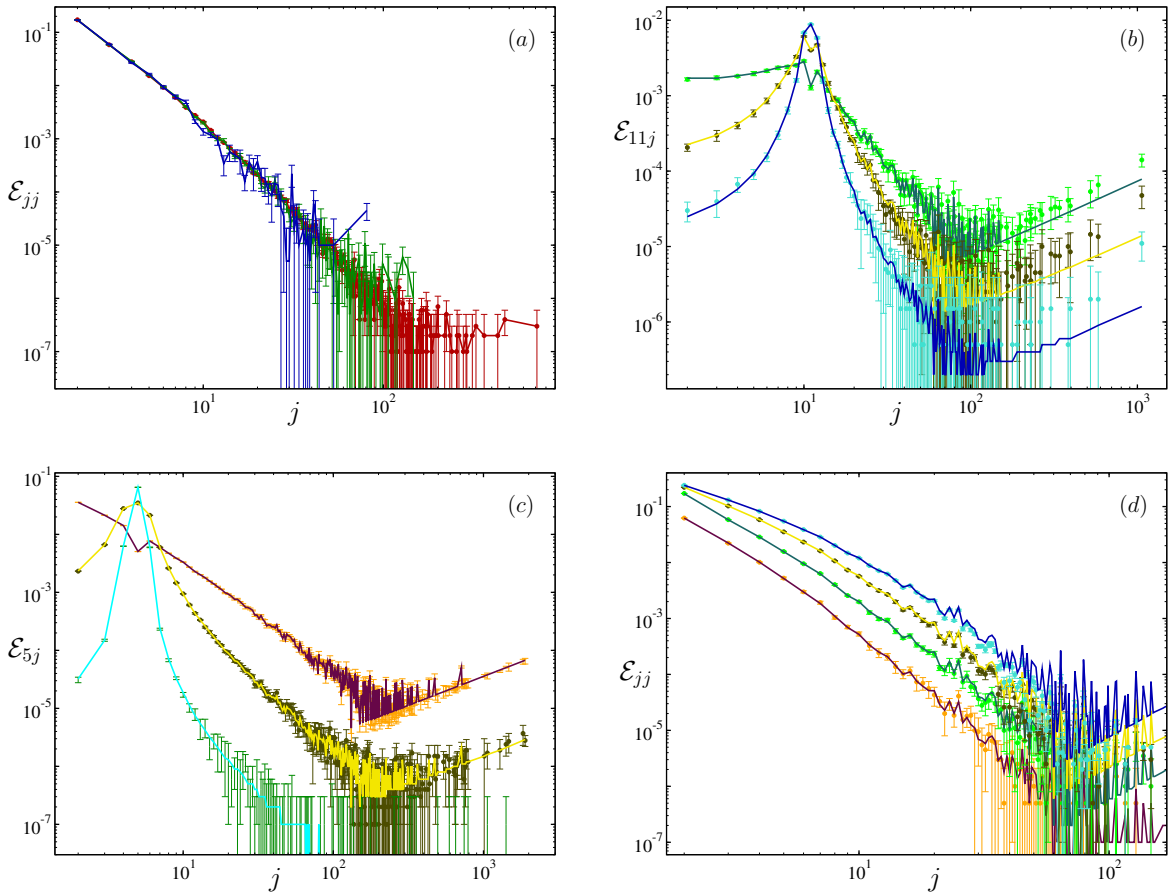

Figure 3.4: Panels (a)-(d): Different connection probabilities, $\mathcal{E}_{ij}$, as function of j. See text for more details. The points with the corresponding error bars represent the measured $\mathcal{E}_{ij}$ of the different assortative networks considered. Except for panel a, the curves correspond to the theory, Eq. 3.21. Note the excellent agreement.

the behavior $\mathcal{E}_{20j}/\mathcal{E}^r_{20j} \geq 1$ for large j and slightly assortative networks shown in the panel.

To summarize, figure 3.3 shows that the AS algorithm is effectively capable of generating assortatively mixed networks. Now, in order to assess the goodness of Eq. 3.21, we plot in figure 3.4 the measured connection probabilities $\mathcal{E}_{ij}$ of some assortative networks generated using the AS algorithm and then compare them with the corresponding theoretical values $\mathcal{E}_{ij}$. All generated networks are based on scale-free Barabási-Albert networks ($m = 2$) of different orders. To calculate the theoretical probabilities, Eq. 3.21, we use the particular degree distribution $P(k)$ of the Barabási-Albert network considered. Panel (a) shows $\mathcal{E}_{jj}$ as function of j for three differently sized networks: $N = 10^4$ (blue), $N = 10^5$ (green), and $N = 10^6$ (rot). The curves correspond to uncorrelated versions ($p = 0$). The picture only demonstrates that our program works satisfactorily; in the limit case of $p = 0$ the resulting networks are actually uncorrelated (compare the picture with figure 3.2, panel a). Note also that according to our theoretical development, the generated correlations depend only on the parameter p and the given degree sequence, but not on the order of the network. Panel (b) shows the probabilities of connection $\mathcal{E}_{11j}$ as function of j for three different assortative versions of a network with $N = 10^5$ vertices. The degree of assortativity is, from top to bottom, $\mathcal{A} = 0.221$, $\mathcal{A} = 0.443$, and $\mathcal{A} = 0.640$.

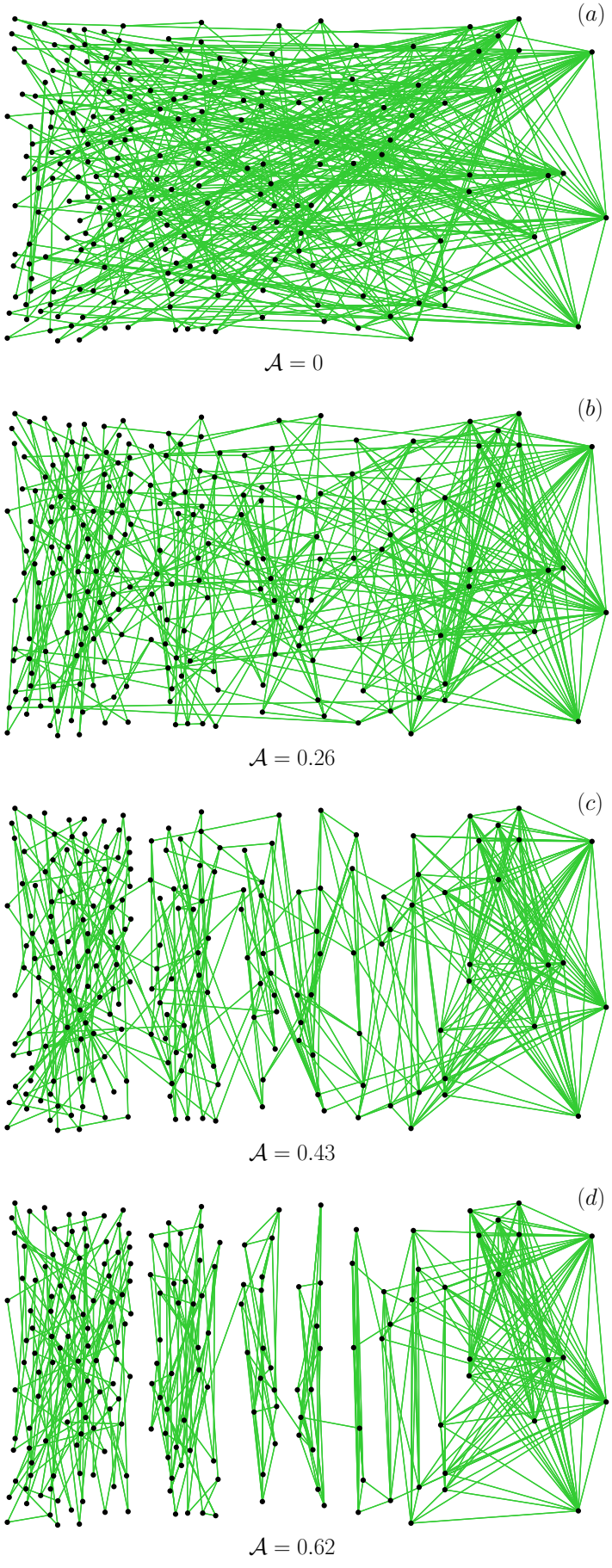

Figure 3.5: Four versions, each one with different degree of assortativity, of a small scale-free network with $N = 200$ vertices and $L = 400$ edges (see text for details). Vertices of same degree are grouped together so that the degree is non-decreasing from left to right. The picture shows (a): $\mathcal{A} = 0$ (uncorrelated network), (b): $\mathcal{A} = 0.26$, (c): $\mathcal{A} = 0.43$, (d): $\mathcal{A} = 0.62$ (maximal assortativity), from top to bottom.

The points with error bars correspond to the measured $\mathcal{E}_{11j}$ and the curves to the theoretical predictions of Eq. 3.21. In panel (c) we compare the measured $\mathcal{E}_{5j}$ (points) with the theory (curves). The simulation results correspond to different assortative versions ($\mathcal{A} = 0$, $\mathcal{A} = 0.443$, and $\mathcal{A} = 0.777$, from top to bottom) of the same original network (in this case with $N = 10^6$ vertices). Finally, panel (d) represents $\mathcal{E}_{jj}$ as function of j for generated versions having the following degree of assortativity: $\mathcal{A} = 0$, $\mathcal{A} = 0.221$, $\mathcal{A} = 0.443$, and $\mathcal{A} = 0.640$ (from bottom to top). All curves correspond to assortative correlated versions of an initial network with $N = 10^5$ vertices. Again points correspond to the measured $\mathcal{E}_{jj}$ and the corresponding curves to the theory. We note that agreement between simulations and the Eq. 3.21 is excellent.

Finally, we conclude the study of the properties of the AS algorithm by drawing a small network to show graphically how the algorithm works. We start again from an initial Barabási-Albert scale-free network, but now with only $N = 200$ vertices and $L = 400$ edges (as above, $m = 2$). To obtain other networks with exactly the same degree sequence but different degree of assortativity, we apply the algorithm discussed. Figure 3.5 shows the changes in the network with varying parameter p. In the figure we have placed the vertices in such a way that vertices of degree 2 are shown in the left part of each panel, all vertices of degree 3 lie to the right of any vertex of degree 2, and all vertices of degree 4 to the right of any vertex of degree 3, etc. The vertices of the same degree are randomly spread within the corresponding area of the figure to better show the edges. The network of maximal accessible assortativity is shown in panel (d). In this network almost all vertices with the same degree are linked between themselves only. The panel shows that all vertices of degree 2 form separated clusters (a more careful analysis unveils that there are three "pearl necklace" clusters of $N = 23$, $N = 30$, and $N = 48$ nodes). All vertices of degree 3 are linked between themselves except for one, which is linked with a node of degree 4. Note that since there are $N_3 = 41$ vertices of degree 3 in our small network, their edges cannot be redistributed within the set. If this was possible, the overall number of edges would be $41 \times 3/2 = 61.5$, since each node bears 3 edges and each of these edges is counted twice in the set. All vertices of degree 4 form a single cluster, with two outgoing links, one to the cluster composed of vertices of degree 3, and one to the cluster of vertices of degree 5. In fact, the network is not a set of isolated clusters of vertices with the same degree only because of the restrictions imposed by the given degree sequence. This restrictions are also responsible for the fact that $\mathcal{A} < 1$, (in this example-network, the maximal assortativity is $\mathcal{A}_{\max} = 0.62$). The other panels show the aspect of the network for diverse values of $\mathcal{A}$.

The dissortative model. A minor change in the AS algorithm can produce dissortative mixing too. We start from a network with a given degree sequence. At each step, we choose randomly two edges of the network and, again, we order the four corresponding vertices with respect to their degrees. Now, however, the edges are rewired in such a way that: With probability p, one edge connects the highest-degree and lowest-degree vertices to each other, and the other edge connects the two remaining vertices; With probability $1 - p$ we rewire the edges randomly. In case the switching process generates a loop or a multiedge, the step is discarded and a new pair of edges selected. The algorithm only depends on the parameter p, and obviously preserves the degree sequence of the network. We will see that, by varying the parameter p, it is possible to construct networks with different degrees of dissortativity, ranging from totally dissortative ($p = 1$) to fully random ($p = 0$).

The corresponding connection probabilities $\mathcal{E}^d_{ij}$ of this model can also be calculated. We consider again the change of variable F_{ln} (Eq. 3.13) when the rewiring procedure is carried out. The analysis of the algorithm reveals that, every time the DS switching step is applied, F_{ln}, either increases or decreases by unity, or does not change. The process corresponds again to an ergodic Markov chain, which ensures the existence of a stationary solution. As in the assortative

case, the effect of loops and multiedges can also be disregarded, since they are rare in the limit of large networks. The solution will thus be exact for psedographs, but, as before, it will be a very good approximation for large simple graphs. Taking all possibilities of change into account, we get the following probabilities of change:

$$\Pi_{+1}^{d} = 2\left[(X_{ln} - f_{ln})^2 + p\left(X_{ln}^2 + (f_{1n} - f_{1,l-1})(f_{1n} - f_{1,l-1} - 2X_{ln})\right)\right] \qquad F_{ln} \to F_{ln} + 1 \quad (3.23)$$

$$\Pi_{-1}^{d} = 2f_{ln}\left[(1 - 2X_{ln} + f_{ln}) + p(f_{1,l-1} + f_{1n} - f_{ln} - 2X_{1,l-1})\right] \qquad F_{ln} \to F_{ln} - 1 \ , \quad (3.24)$$

where we have already approached $X_{ln} - f_{ln}$ by $X_{ln} - f_{ln} - 1$. Here, $f_{ln} = F_{ln}/L$, and X_{ln} is given by Eq. 3.16. The expected values $\langle f_{ln} \rangle := \mathcal{F}_{ln}^{d}$ $(l < n)$, are then given in this case by the conditions:

$$\left(X_{ln} - \mathcal{F}_{ln}^{d}\right)^2 - p\left(X_{ln}^2 + (\mathcal{F}_{1n}^{d} - \mathcal{F}_{1,l-1}^{d})(\mathcal{F}_{1n}^{d} - \mathcal{F}_{1,l-1}^{d} - 2X_{ln})\right) = \quad (3.25)$$

$$= \mathcal{F}_{ln}^{d}\left((1 - 2X_{ln} + \mathcal{F}_{ln}^{d}) + p(\mathcal{F}_{1,l-1}^{d} + \mathcal{F}_{1n}^{d} - \mathcal{F}_{ln}^{d} - 2X_{1,l-1})\right) \ ,$$

for all $l > 1$, and

$$(1-p)\left(X_{1n} - \mathcal{F}_{1n}^{d}\right)^2 = \mathcal{F}_{1n}^{d}\left(1 - 2X_{1n} + \mathcal{F}_{1n}^{d}\right) \ , \quad (3.26)$$

for $l = 1$. Note that Eq. 3.25 reduces to Eq. 3.26 when $l = 1$ (given that X_{ln} and f_{ln} vanish when one of the indices is smaller than 1, the minimal tolerated degree). Using Eqs. 3.25 and 3.26, we obtain after some calculations:

$$\mathcal{F}_{ln}^{d}(p) = \begin{cases} \dfrac{1}{p}\left[\mathcal{C}_{ln} - \sqrt{(\mathcal{C}_{ln})^2 + (D_n - D_{l-1})^2 - pX_{ln}^2}\right] & \text{if} \quad 0 < p \leq 1 \\ X_{ln}^2 & \text{if} \quad p = 0 \ . \end{cases} \quad (3.27)$$

where $\mathcal{C}_{ln}$ and D_k respectively are:

$$\mathcal{C}_{ln} = \frac{D_n + D_{l-1} + pX_{ln}}{2} \qquad \text{and} \qquad D_k = \sqrt{\left[pX_{1k} - \frac{1}{2}\right]^2 + p(1-p)X_{1k}^2} \ . \quad (3.28)$$

Finally, to obtain the degree-degree correlations corresponding to the model, we come back to the definition of F_{ln}, Eq. 3.13, and find $\mathcal{E}_{ij}$:

$$\mathcal{E}_{ij}^{d}(p) = \mathcal{F}_{ij}^{d}(p) - \mathcal{F}_{i,j-1}^{d}(p) - \mathcal{F}_{i+1,j}^{d}(p) + \mathcal{F}_{i+1,j-1}^{d}(p) \ . \quad (3.29)$$

One can prove that Eq. 3.29 reduces to the uncorrelated solution $\mathcal{E}_{ij}$ when $p = 0$. For $p = 1$ the solution vanishes completely for a lot of values of indices i and j. An accurate analysis of case $p = 1$ shows that vertices of low degree connect only to vertices of large degree, and vice versa.

Dissortativity means that high-degree vertices tend to connect to low-degree vertices with larger probability than in an uncorrelated network. For undirected networks, this means that low-degree vertices connect preferably to high-degree vertices too, and consequently, that vertices of moderate degrees tend to connect among themselves. In highly dissortative networks this tendency is very strong. Let us assume we are constructing a fully dissortative scale-scale network. In an intuitive way, we should connect all vertices having the maximum degree to vertices of minimum degree. Once all vertices having the maximum degree are exhausted, the vertices of the second highest degree should also be connected to vertices having the minimum degree (this is always possible, since in a scale-free network low-degree vertices build an overwhelming majority). Also vertices of the third, fourth, etc., highest degree might be connected

with vertices having the smallest degree, until all vertices of minimum degree are connected. After this, vertices having the second minimum degree should be connected to those vertices having the highest degree not yet connected, and so on. Our model generates exactly this type of correlations for $p = 1$; both simulations and theory (Eq. 3.29) confirm this scenario.

Let us analyze the correlations appearing numerically. To this end, we carry out extensive numerical simulations based on a Barabási-Albert network with $N = 3 \cdot 10^4$ vertices and double number of edges ($m = 2$). All points are averaged again over ten independent realizations of the algorithm when applied to the original network. The outcomes are summarized in picture 3.6. Panel (a) shows how the nearest neighbors' average degree function, $\bar{k}_{nn}(j)$, varies when the parameter p changes. (We use here p as a measure of the dissortative mixing in the networks. Although the parameter p is an internal parameter of the algorithm and does not immediately represent any property of the network, it is clear that any reasonably defined degree of dissortativity has to be an increasing function of p.) The curves correspond, from top to bottom, to $p = 0$, $p = 0.5$, $p = 0.78$, $p = 0.9$, $p = 0.999$, and $p = 1$. We see that $\bar{k}_{nn}$ remains constant (upper curve) and equal to $\bar{k}_{nn} = \langle k^2 \rangle / \langle k \rangle$ (horizontal straight line) when the network is uncorrelated ($p = 0$). However, when correlations are modified by the increasing of parameter p, this flat curve is transformed into a decreasing one, indicating the appearance of dissortative mixing; when increasing p, $\bar{k}_{nn}(j)$ also increases for j small, and decreases for $j \gg 1$. This is exactly the typical behavior what is expected for dissortative networks. Consider now the case $p = 1$. We expect the existence of a "peak" at j_{min} (the minimum degree of the network, $j_{min} = 2$ in our case) and a plateau at $\bar{k}_{nn} = j_{min}$ for $j \gg 1$, since in this case all hubs must be linked to vertices of minimum degree, and vice versa. Both features are revealed in our simulations (see lower curve of the picture).

The assertion that highly dissortative networks exhibit assortative mixing among moderate-degree vertices is also confirmed by the simulations. A careful inspection of the numerical results certainly shows that medium-degree vertices tend to connect between themselves as the dissortativity increases. Panels (b), (d), and (f) of figure 3.6 display, respectively, $\mathcal{E}_{2j}/\mathcal{E}^r_{2j}$, $\mathcal{E}_{3j}/\mathcal{E}^r_{3j}$, and $\mathcal{E}_{4j}/\mathcal{E}^r_{4j}$ as function of j. In each panel, the following values of p are considered: $p = 0$ (maroon), $p = 0.5$ (green), $p = 0.78$ (orange), $p = 0.9$ (yellow), $p = 0.99$ (blue), and $p = 1$ (black). Let us primarily focus on panel (b). We observe that vertices of degree $j_{min} = 2$ are more likely to connect to high-degree vertices as the dissortativity increases. In the extreme case $p = 1$, vertices of degree 2 connect only to vertices of degree 10, or larger; and vice versa, all vertices of degree larger than 10 connect only to vertices of degree 2 (vertices of degree 10 connect only to vertices of degree 2 or 3). The peculiar form of $\mathcal{E}_{2j}/\mathcal{E}^r_{2j}$ for $p = 1$ can also be explained. If all vertices with a given degree j connect to vertices of degree 2, then $\mathcal{E}_{2j} = 2jP(j)/\langle j \rangle$, and consequently

$$\frac{\mathcal{E}_{2j}}{\mathcal{E}^r_{2j}} = \frac{\dfrac{2jPj}{\langle j \rangle}}{(2 - \delta_{2j})\dfrac{2P(2)jP(j)}{\langle j \rangle^2}} = \frac{\langle j \rangle}{2P(2)} = 4 \tag{3.30}$$

(for all $j \neq 2$). We expect then that all values of $\mathcal{E}_{2j}/\mathcal{E}^r_{2j}$ are very close to 4 (exactly 4 won't work because hubs do not correspond to Eq. 3.1). Panel (b) shows exactly this. Now, if all vertices of degree larger than 10 link to vertices of degree 2, vertices of degree 3 must necessarily connect to vertices of degree 10 or smaller. This is exactly what we can see in panel (d): for $p = 1$, all vertices of degree 3 connect to vertices of degrees between $j = 6$ and $j = 10$, and all vertices of degree 7, 8, and 9 connect to vertices of degree 3 (vertices of degree 6 connect only to vertices of degree 3 or 4). Proceeding in a similar way as in Eq. 3.30, one can easily prove that $\mathcal{E}_{3j}/\mathcal{E}^r_{3j} = 20/3$, which is confirmed by the simulations (the value 20/3 is indicated

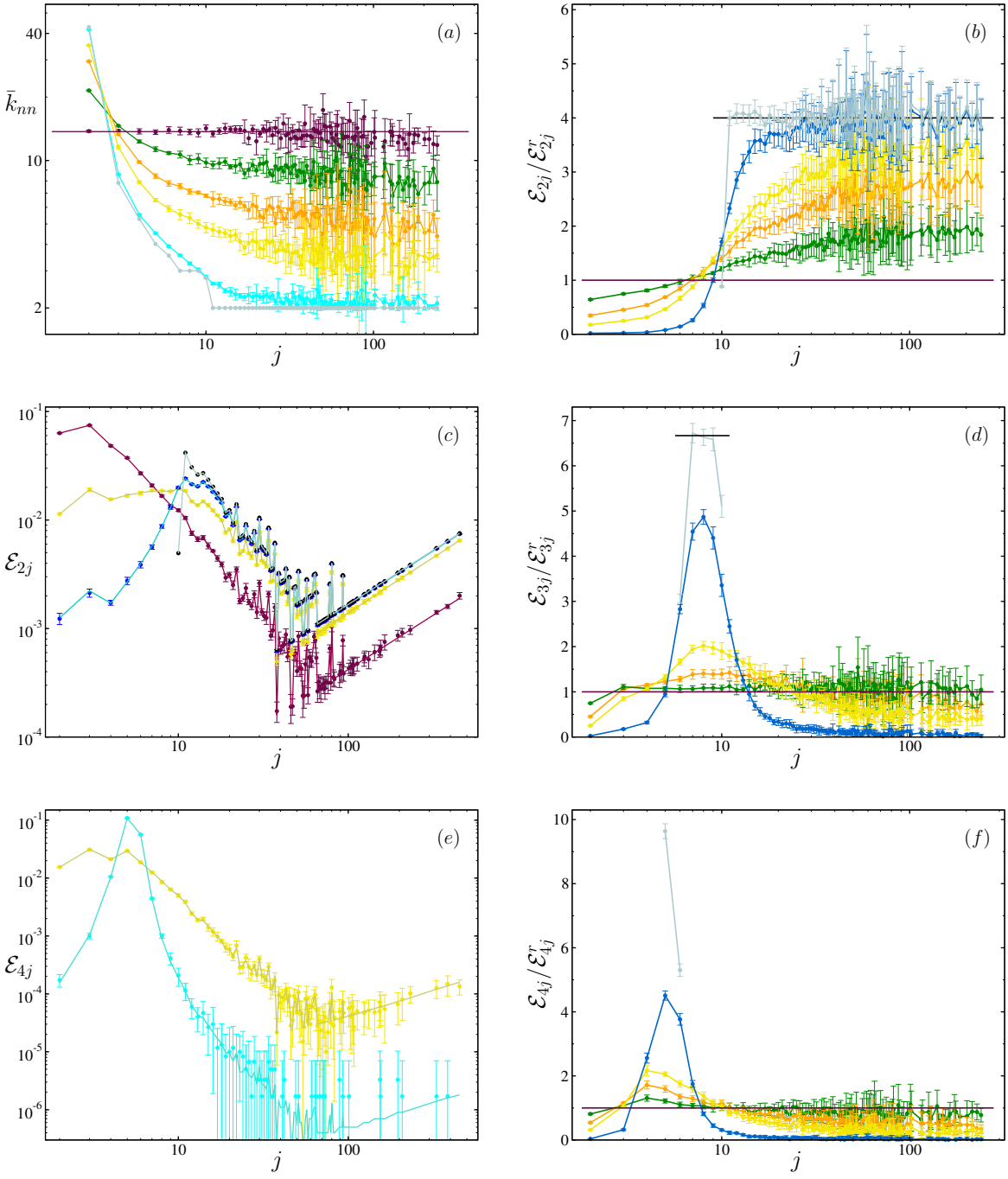

Figure 3.6: Panel (a): Nearest neighbors' average degree $\bar{k}_{nn}(j)$ as function of j for different dissortative versions of a Barabási-Albert network. Using the parameter p to measure dissortativity, the panel shows, from bottom to top, $p = 0$ (uncorrelated), $p = 0.5$, $p = 0.78$, $p = 0.9$, $p = 0.999$, and $p = 1$. The horizontal straight line corresponds to $\bar{k}_{nn}(j) = \langle k^2\rangle/\langle k\rangle$. Panels (c) and (e): $\mathcal{E}_{2j}$ and $\mathcal{E}_{4j}$, respectively, as function of j. Here the results correspond to networks with $p = 0$ (maroon), $p = 0.9$ (yellow), 0.99 (blue), and $p = 1$ (grey), in panel (c), $p = 0.9$ (yellow), and $p = 0.999$ (cyan), in panel (e). In both pictures the points correspond to the simulations and the curves to the corresponding theoretical results (Eq. 3.29). Panels (b), (d), and (f): $\mathcal{E}_{2j}/\mathcal{E}^r_{2j}$, $\mathcal{E}_{3j}/\mathcal{E}^r_{3j}$, and $\mathcal{E}_{4j}/\mathcal{E}^r_{4j}$ as function of j. See text for more details.

by the small black straight line in the picture). The study of case $j = 4$ (panel f) shows a similar behavior: All vertices of degree 4 connect only to vertices of degrees 5 or 6 when $p = 1$. For smaller p, vertices of degree 4 may connect to vertices of different degrees than 5 or 6, but the probability that they connect to a vertex of degree j decreases rapidly when j is $j \gg 5$ or $j \ll 5$. This is an explicit indication of assortative behavior among moderate-degree vertices (in this case, among vertices of degree close to 5), to the effect that similar degree vertices connect between themselves). Furthermore, the more dissortative the network the stronger this assortative mixing is.

All these peculiarities of the dissortative networks are in perfect agreement with the theory. In figures 3.6 (c) and (e) we compare the simulation results of some $\mathcal{E}_{ij}$ with the theoretical outcomes of Eqs. 3.27 and 3.29. (Here, the simulation results are based on a Barabási-Albert network with a different degree sequence than the network of the previous results). Figure 3.6 (c) shows $\mathcal{E}_{2j}$ as function of j for the cases $p = 0$ (maroon), $p = 0.9$ (yellow), $p = 0.99$ (blue), and $p = 1$ (grey). Figure 3.6 (e) shows the behavior of $\mathcal{E}_{4j}$ for the cases $p = 0.9$ (yellow), and $p = 0.999$ (cyan). In both pictures, dots (with error bars) correspond to the numerical findings, while curves represent the corresponding theoretical results. We note excellent agreement. We also tested the theory using other $\mathcal{E}_{ij}$ and found the same excellent accordance.

To conclude the study of dissortative networks, we discuss the problem of the measure of dissortativity. Finding a suitable parameter to evaluate the degree of dissortativity of a network is not an easy task. The best proof of this is the fact that strong dissortativity implies strong partial assortativity. To attack the question, let us come back to our theoretical results, Eq. 3.27. For $p = 1$, this equation vanishes completely in two cases: i) when $l = 1$ and n satisfies $2X_{1n} \leq 1$, (that is, no vertices of degree n or lower are connected to each other), and ii) when $l = q + 1$, where q is any integer for which $2X_{1q} \geq 1$, and $n = K$, the maximum degree of the network (no edges connect vertices of larger degree than $q + 1$ to each other). Case i): When $l = 1$ Eq. 3.27 reduces to

$$\mathcal{F}^{d}_{1n}(p) = \frac{1}{p}\left[pX_{1n} - \frac{1}{2} + \sqrt{\left(pX_{1n} - \frac{1}{2}\right)^2 + p(1-p)X_{1n}^2}\right] \tag{3.31}$$

(expression which can also be obtained from Eq. 3.26). Making $p = 1$ in this last equation we have $\mathcal{F}^{d}_{1n}(p = 1) = X_{1n} - 1/2 + |X_{1n} - 1/2|$, which obviously vanishes if $2X_{1n} \leq 1$. Case ii) Eq. 3.27 takes the form

$$\mathcal{F}^{d}_{q+1,K} = \frac{\left|X_{1,K} - \frac{1}{2}\right| + \left|X_{1,q} - \frac{1}{2}\right| + X_{q+1,K}}{2} - \tag{3.32}$$

$$- \sqrt{\left(\frac{\left|X_{1,K} - \frac{1}{2}\right| + \left|X_{1,q} - \frac{1}{2}\right| + X_{q+1,K}}{2}\right)^2 + \left(\left|X_{1,K} - \frac{1}{2}\right| - \left|X_{1,q} - \frac{1}{2}\right|\right)^2 - X_{q+1,K}^2}$$

for $p = 1$. Now, we impose the condition $2X_{1q} \geq 1$. Taking into account that $X_{1,K} = 1$ and $X_{1,q} + X_{q+1,K} = 1$, Eq. 3.32 reduces to

$$\mathcal{F}^{d}_{q+1,K} = \frac{1}{2} - \sqrt{\frac{1}{2} + (1 - X_{1,q})^2 - X_{q+1,K}^2} = 0 \ . \tag{3.33}$$

This proves that only edges connecting large- to low-degree vertices can be found in the maximally dissortative case. Inspired by these results, we would like to propose the following quantity

as coefficient of dissortativity:

$$\mathfrak{D} = \frac{(\mathcal{F}_{1\overline{k}} + \mathcal{F}_{\underline{k},K}) - (\mathcal{F}^{r}_{1\overline{k}} + \mathcal{F}^{r}_{\underline{k},K})}{(\mathcal{F}^{r}_{1\overline{k}} + \mathcal{F}^{r}_{\underline{k},K})} \tag{3.34}$$

where $\overline{k}$ is the maximum integer for which $2X_{1\overline{k}} \leq 1$, $\underline{k}$ is the minimum integer for which $2X_{1\overline{k}} \geq 1$, $K =$ the maximum degree of the network, and the superscript r indicates that the values of $\mathcal{F}_{ln}$ must be calculated when the network is uncorrelated. This coefficient takes the value -1 when the network is fully dissortative and vanishes when the network is uncorrelated.

Hence, simulations and theory demonstrate that the DS algorithm generates dissortative mixing. For $p = 1$ the algorithm produces totally dissortative networks, and for $p = 0$ random mixing. Additionally, the model makes no distinction between vertices of same degree, that is, the connections between same degree vertices are random. This fact allows us to state that the algorithm creates "random dissortative" correlations, i. e., that it produces dissortative mixing leaving random all other properties of the network. Although the model is a further step in understanding dissortative correlated networks, we want to remark that the connections E^d_{ij} which it generates may not be considered the only ones capable of satisfying the condition: vertices of large degree connect preferably to low-degree vertices. The same remark is valid for our AS model: the AS algorithm is certainly not the only link-restructuring process which is able to increase the number of connections between similar degree vertices (with respect to the uncorrelated network). In effect, many real networks show assortativity or dissortativity, and, at the same time, other properties that our models do not reproduce. Thus, the Internet shows $\bar{k}_{nn} \propto j^{-\nu}$ [Vázquez et al., 2002], but this feature is not viewable in figures 3.6 (a) and 3.2 (c). This is not surprising since real-world networks are also governed by the metrics of the underlying space, normally the Euclidean physical space, an ingredient that we did not introduce in previous algorithms. In spite of this, the introduced models are very interesting from a mathematical viewpoint, and exhibit very remarkable properties which mimic those of real networks. We discuss them in the last section of the chapter.

3.3 Highly triangulated scale-free networks

It is found in many networks that if vertex i is connected to vertex j and vertex j to vertex l, then there is a heightened probability that vertex i will also be connected to vertex l. *Transitivity* and *clustering* are the terms currently used in network topology to refer to this probability[5]. The notion has its roots in sociology, where it was important to analyze the groups of acquaintances in which every member knows every other one. In the language of social networks, the friend of your friend is likely also to be your friend. Although diverse clustering coefficients are mentioned in corresponding literature [Dorogovtsev, 2004, Soffer and Vázquez, 2005], transitivity is usually quantified by means of the mean clustering coefficient, C (Eq. 1.1), which is basically a measure of the number of cycles of length 3 in the network. The degree-dependent clustering coefficient $C(k)$ is also used very often, particularly since it was found that many real networks show the power law behavior $C(k) \sim k^{-\gamma}$. Transitivity plays an essential role in the description of processes taking place in networks, like spreading phenomena [Sander et al., 2002, Hufnagel et al., 2004, Eguíluz and Klemm, 2002, Pastor-Satorras and Vespignani, 2001], or in percolation [Cohen et al., 2000a].

[5]Although both terms are indiscriminately used in scientific literature, we prefer, like [Newman, 2003b], to use the term "transitivity" to designate the probability described above, and relegate the term "clustering" to characterize the tendency of vertices of many real networks to highly connected group forming clusters (community structure).

Networks in the real world exhibit large degrees of transitivity, typically of order $C \sim 0.1$. However, the network models we have nowadays do not attain to reproduce successfully this characteristic. Thus, small world or lattice models present large transitivity degrees, but they cannot reproduce other important features like the scale-free character of networks. On the other hand, patterns of evolving scale-free networks usually show very small mean clustering coefficients, which vanish when the order of networks diverges. The first model capable of coordinating arbitrary degree distributions and large degrees of transitivity was proposed by Volz [2004]. In his interesting work Volz demonstrates that the wiring structure of connections changes considerably in networks with Poisson and exponential degree distributions as the transitivity increases. The model, however, presumably does not work perfectly on scale-free networks, because it cannot always successfully operate in the presence of hubs. In this section we propose another model capable of combining both ingredients, arbitrary degree distributions and transitivities, in the most general way. The model, based like the preceding algorithms on edge-restructuring processes, can successfully change transitivity of scale-free networks to a desired degree without modifying the degree sequence of networks.

The algorithm we suggest here was motivated by the following problem: Find how the connections between vertices can be restructured in a given network in order to generate "maximal triangulation", preserving the original degree sequence of the network. Here, a network is said to be maximally triangulated when the mean clustering coefficient takes the maximum possible value C_{max}. We note that a network can have a mean clustering coefficient equal to unity, $C = 1$, if and only if it consists of isolated cliques, each one composed of $k+1$ vertices of degree k. This means that the value $C = 1$ is only possible if the number of vertices of degree k, N_k, is such that, for each k, N_k is multiple of $k + 1$; in any other case the mean clustering coefficient will be strictly smaller than one. This reasoning must be sufficient to convince us that most networks can never reach the ideal value $C = 1$; at most a value $C = C_{max} < 1$. This latter is particularly true for finite scale-free networks, whose principal characteristic is that $N_k \ll k$ for large k (hubs).

We can construct networks (almost) maximally triangulated as follows: i) We label the vertices of the network by parting at random the vertices of the network into $\mathrm{int}[N_k/(k+1)]$ subsets of $k+1$ vertices of degree k (for all k) and one additional subset containing the rest of the vertices $(\sum_k N_k - \mathrm{int}[N_k/(k+1)](k+1))$. Here, $\mathrm{int}[x]$ represents the integer part of x. We then assign a label to all subsets thus formed. The label of each set will be used as a label for all vertices belonging to the set. For notation purposes, each subset composed of $k+1$

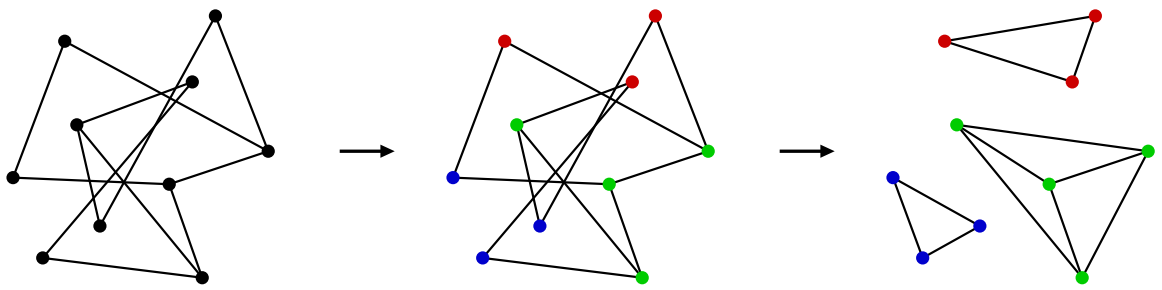

Figure 3.7: Sketch of the functioning of the TS model. The algorithm starts from a network with a given degree sequence. Then, a label is assigned to each vertex (each label is represented by a color drawing). Finally, the TS rewiring process is implemented. The resulting network finally consists of a set of cliques ($p = 1$). See text for details.

vertices of degree k is said to be a k-subset, and the additional subset a remaining-subset. ii) We rewire the connections between vertices. At each step of the rewiring process, we choose at random two edges of the network. If two of the four end-vertices associated with the selected links are the same vertex, we repeat the procedure and choose a new pair of edges. If the four end-vertices are different, we normally have four vertices bearing different labels. Then, if two and no more then two of these vertices bear the same label, the edges are rewired in such a way that one edge connects the two vertices with the same label and the other edge connects the remaining vertices; otherwise, the edges are randomly rewired. In case one or both of these new edges already exists in the network, this step is discarded and a new pair of edges is selected. As explained above, the latter restriction prevents as above the appearance of multiple edges connecting the same pair of vertices. The repeated application of the rewiring step leads to a very distinct triangulated version of the original network. Note that the larger the network, the closer to unity the value of the mean clustering coefficient will be. Finally, note that the algorithm does not change the degree of vertices, and therefore, it preserves the initial degree sequence of the network.

Now, to construct highly but not fully triangulated networks, we proceed as follows: Once we have already labeled the vertices, we apply the above explained rewiring process with probability p only, while with probability $1-p$ we rewire the links at random. Changing the parameter p it is possible to construct networks with arbitrary mean clustering coefficient, ranging from the C corresponding to uncorrelated networks ($p = 0$) to the corresponding to maximally triangulated ($p = 1$). Note that the process of repeated application of the algorithm corresponds to an ergodic Markov chain, and therefore the system will always reach a stationary state [Xulvi-Brunet and Sokolov, 2004]. This general algorithm, which we will call "transitive switching" (TS) algorithm, has the effect that vertices try to form cliques in which all vertices are interconnected with each other[6], under the restrictions imposed by the preassigned degree sequence of the network. This tendency is opposed by random mixing.

We apply the TS algorithm to Barabási-Albert networks using different values of parameter p in order to study the statistical properties of highly triangulated scale-free networks. The network to which the algorithm is applied is a Barabási-Albert network of $3 \cdot 10^4$ vertices and $6 \cdot 10^4$ edges, which again corresponds again to $m = 2$ attached edges per step. The simulation results are averaged over ten realizations of the algorithm. Figure 3.15 (a) shows how the fraction of vertices, $\mathcal{M}$, of the giant component changes when the mean clustering coefficient increases. We see that the network remains connected until the mean clustering coefficient reaches the value $C \simeq 0.12$ -two orders of magnitude larger than in the uncorrelated case). From this value on, the network tends to separate in distinct disconnected clusters and the mass of the giant component decreases[7]. When transitivity approaches to its maximum value ($C_{max} \simeq 0.995 \sim 1$ for our simulations), the network breaks down into many small, fully connected cliques, with each vertex in a clique sharing a common degree. The behavior of the average path length, l, (panel b of the figure) will be commented on in greater length in the next section. Our simulations also show that when increasing C, the clusters which first separate from the giant component are typically triples of connected nodes, and, in general, that the order of these separated clusters grows slowly when C increases. An interesting consequence of this behavior is that the vertex degree distribution of the giant cluster, $P_c(k)$, goes on exhibiting a power law tail until the system is highly triangulated (figure 3.8 a).

The study of nearest neighbors' average degree function, $\bar{k}_{nn}(k)$, shows also interesting

[6] In sociological terms, this would be the equivalent of persons trying to form groups in which every member knows every other member.

[7] This behavior had already been observed in networks with large degrees of transitivity [Newman et al., 2001, Volz, 2004].

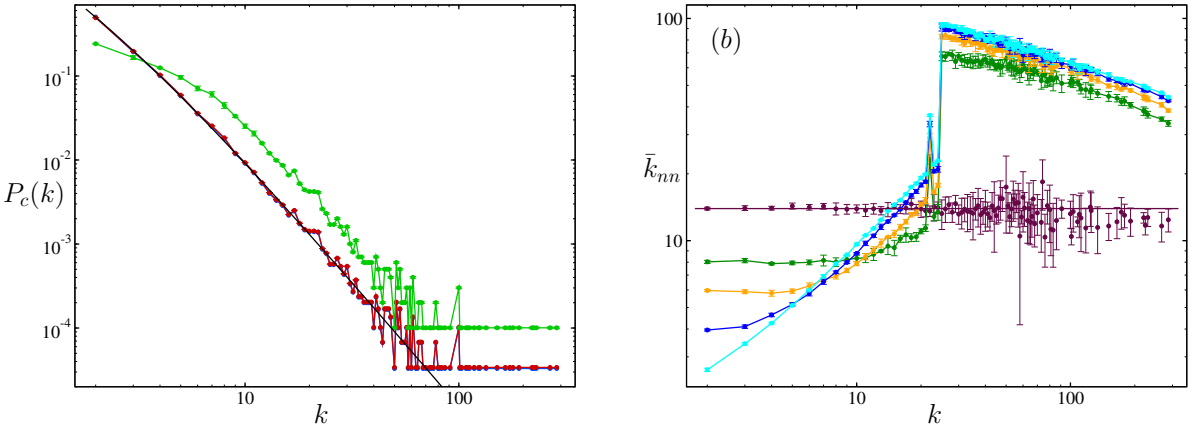

Figure 3.8: Panel (a). Degree distribution, $P_c(k)$, of the giant component as function of k. The curves correspond to: $C = 0.0013$, the uncorrelated case (blue), $C = 0.1202$ (red), $C = 0.598$ (green), and Eq. 1.6 (black). Panel (b). Nearest neighbors' average degree, $\bar{k}_{nn}(k)$, versus k for the TS networks which have the following degree of transitivity: $p = 0$ (maroon straight line), $C = 0.0025$ (dark green), $C = 0.0126$ (orange), $C = 0.1203$ (blue), and $C = 0.598$ (cyan).

degree-degree correlations for the TS networks. Panel (b) of figure 3.8 shows $\bar{k}_{nn}(k)$ as a function of j for the same networks of panel (a). First of all we want to comment on the discontinuity that the picture shows at $j = 24$. It appears because the original Barabási-Albert network to which the TS algorithm is applied has a degree sequence with $N_k > k$, $\forall k \leq 24$ and $N_k < k$, $\forall k > 24$. Therefore, the discontinuity arises as a consequence of the algorithm itself, and it is due to the finite order of the network. In the limit of large orders, $N \to \infty$, the discontinuity disappears, and function $\bar{k}_{nn}(k)$ shows a behavior as exhibited in the figure for values of j smaller than 25. Note that the fraction of vertices of degree larger than $k = 24$ is small, and therefore, not decisive for the global behavior. Without taking the finite size effects into account, the simulation results indicate that the assortativity grows when the transitivity of the network is increased. The interesting feature of these strongly triangulated networks is that assortative mixing appears primarily between large degree vertices in slightly triangulated ones, and that, when increasing the degree of transitivity, low-degree vertices connect more and more assortatively.

Let us compare this model with the AS algorithm. Both algorithms operate in quite a similar way, only that the AS model considers the degree of vertices as "input" during the switching process, while the TS model takes into account the labels assigned to the vertices. The small differences between both rewiring processes are however fundamental when it comes to reproducing the appropriate connection features of the models. For example, the AS algorithm connects vertices of degree i to vertices of degree $j \simeq i$ with a larger probability than to vertices of degrees $j \gg i$ or $j \ll i$. If the AS switching step was used to rewire the edges in the TS model, then l-labeled nodes would preferably connect to similar-labeled vertices, and connect to vertices with labels either much larger or much smaller than l with less probability; given a l-subset, for example the l-subset a, and a certain number n of k-subsets, subset a would connect to each k-subset with a different probability only because of the fact that the k-subsets have a different label. Consequently, vertex-pair correlations would then depend on how the process of assignment of labels is carried out. The TS algorithm, however, works in such a way that, regardless of the

label that they have assigned, all k-subsets exhibit identical connection probabilities when the stationary state is reached. That is, two k-subsets could be interchangeabled without having to modify the correlations. This feature of the TS networks follows immediately from the rewiring procedure implemented in the algorithm: By construction, when two and no more than two of the selected vertices (i. e., the vertices associated to the pair of edges chosen at random) have the same label, the rewiring procedure works regardless of which labels are associated with the remaining vertices, and when either all of the labels of the selected vertices are different or more of two of them are equal, the rewiring procedure is completely random. Numerical simulations confirm that the connections probabilities of the TS model do not depend on the labels assigned to the subsets, but on their structural properties only.

The model produces very interesting vertex-pair correlations. In order to study them analytically, we first focus on how the edges of a given subset, for example m-subset c, remain restructured when the algorithm is applied with a certain value of p. Let U_m^c be the number of edges which connect the vertices of subset c between themselves (internal edges). The number of the edges which connect m-subset c to other subsets of the network (external edges) is then given by $m(m+1) - 2U_m^c$. Analyzing all possibilities of change every time the switching procedure is applied, we can observe that U_m^c either increases or decreases by unity or does not change at all. As a next step, we calculate the corresponding probabilities of change $U_m^c \to U_m^c \pm 1$. This calculation can trivially be carried out if the effects of loops and multiedges are obviated; the probability that U_m^c increases by unity ($U_m^c \to U_m^c + 1$) is given by $(1+p)[m(m+1) - 2U_m^c][m(m+1) - 2U_m^c - 1]/(2L(L-1))$, and that U_m^c decreases by unity ($U_m^c \to U_m^c - 1$) is given by $2(1-p)U_m^c[L - m(m+1) + U_m^c]/(L(L-1))$. Taking both probabilities into account, one can then write the equation which yields the expected fraction of internal edges when the system reaches the stationary state. Unfortunately, the solution that we would get from this equation would only be valid if our network was a pseudograph. In the present case, however, in which loops and multiedges are not allowed, the probability that $U_m^c \to U_m^c + 1$ is in general smaller than $(1+p)[m(m+1) - 2U_m^c][m(m+1) - 2U_m^c - 1]/(2L(L-1))$ due to the fact that, when selected, a non-negligible fraction of the pairs of edges which are counted in $(1+p)[2m(m+1) - 2U_m^c][m(m+1) - 2U_m^c - 1]/(2L(L-1))$, not only increases U_m^c by unity, but also produces loops and multiedges (especially when $p \to 1$). The function that correctly yields the probability that U_m^c increases by unity is

$$\mathcal{P}_{+1}^t(m, U_m^c) \;=\; \frac{1+p}{L(L-1)} \sum_{l=0}^{m-1} \sum_{k=0}^{m-1} \mathbb{P}\left(m, U_m^c; l, k\right) (m-l)(m-k) \;. \tag{3.35}$$

Here, the function $\mathbb{P}(k, n; i, j)$ is defined as

$$\mathbb{P}(k, n; i, j) \;=\; \mathbb{T}(n, i, j)\, \frac{\mathbb{F}(k, i)\; \mathbb{F}(k, j)\; \mathbb{F}(k(k+1)/2 - 2(k+1), n - i - j)}{\mathbb{F}(k(k+1)/2, n+1)} \quad . \tag{3.36}$$

where the functions $\mathbb{T}(n, i, j)$ and $\mathbb{F}(i, j)$ are respectively given by the following expressions

$$\mathbb{T}(n, i, j) = \begin{cases} \dfrac{n!}{i!j!(n-i-j)!} & \text{if} \quad n \geq i + j \\ 0 & \text{if} \quad n < i + j \end{cases} \tag{3.37}$$

and

$$\mathbb{F}(i, j) = \begin{cases} (i-1)(i-2)\cdots(i-j+1)(i-j) & \text{if} \quad j \geq 1 \\ 0 & \text{if} \quad j < 1 \quad . \end{cases} \tag{3.38}$$

The term $\sum_l \sum_k \mathbb{P}(m, U_m^c; l, k)(m-l)(m-k)$ gives the number of pairs of edges which can be selected on average to produce the increase $U_m^c \to U_m^c + 1$ (without creating loops and multiedges). Note that $\sum_l \sum_k \mathbb{P}(m, U_m^c; l, k) = m(m+1)/2 - U_m^c$. On the other hand, the probability that $U_m^c \to U_m^c - 1$ can however be approached by the already given expression $\mathcal{P}_{-1}^t = 2(1-p)U_m^c[L - m(m+1) + U_m^c]/(L(L-1))$, since the probability that multiedges appear in the system connecting vertices of different subsets may be neglected in the limit of large networks. Therefore, the equation which yields the probability $\mathfrak{u}_m^c := \langle U_m^c \rangle / L$ that a randomly selected edge of the network connects two vertices both belonging to the same m-subset is

$$(1+p) \sum_{l=0}^{m-1} \sum_{k=0}^{m-1} \mathbb{P}(m, U_m^c; l, k)(m-l)(m-k) = (1-p)U_m^c \left(1 - m(m+1) + U_m^c\right) \quad . \tag{3.39}$$

Since Eq. 3.39 is not easy to solve analytically, in order to obtain the solution, we make use of numerical methods. We use here the bisection method to solve nonlinear equations described in the numerical recipes [Flannery et al., 1985]. Using this numerical procedure in Eq. 3.39, we can obtain the expected value of U_m^c, and then, using the definition $\mathfrak{u}_m^c := \langle U_m^c \rangle / L$, the probability $\mathfrak{u}_m^c$ that an edge of the network belongs to m-subset c. (For the especial case $p = 0$, for which the probability that a given subset contains internal edges is negligible, Eq. 3.35 may roughly be approached $\mathcal{P}_{+1}^t \simeq (1+p)(m(m+1) - 2U_m^c)^2/(2L(L-1))$. Using this expression in Eq. 3.39, we obtain an equation which can be analytically solved, whose solution is $\mathfrak{u}_m^c = [m(m+1)]^2/4L^2$.)

The validity of Eq. 3.39 is checked in figure 3.9, where we compare the theoretical results given by the equation with numerical simulations. To carry out the comparison, we construct scale-free networks whose degree sequence is such that the partition process into k-subsets can be carried out perfectly, i. e., networks in which the remaining-subset is empty (to eliminate possible biases due to the existence of this set). This condition does not restrict the rank of applicability of this study, since the existence of a nonempty remaining-subset is statistically unimportant in the limit of large networks. We built two scale-free networks, one with $N = 250436$, $L = 334682$, and $P(k) \propto (k+1)k^{-4}$ (network $\mathfrak{a}$), and another one with $N = 39993$, $L = 66439$, and $P(k) \propto (k+1)k^{-3.5}$ (network $\mathfrak{b}$). To both networks the TS algorithm with different values of p is then applied. All simulations results are averaged over ten realizations of the algorithm. Figure 3.9 (a), corresponding to networks of type $\mathfrak{a}$, shows $\mathfrak{u}_m^c$ as a function of $m+1$ (the number of vertices of any m-subset) for the following values of p: $p = 0.9$ ($C = 8.3 \cdot 10^{-6} \pm 5.6 \cdot 10^{-6}$), $p = 0.999$ ($C = 7.9 \cdot 10^{-5} \pm 1.1 \cdot 10^{-5}$), $p = 0.9999$ ($C = 0.00413 \pm 0.00004$), and $p = 0.99999$ ($C = 0.0677 \pm 0.0003$). The mean clustering coefficient C of the uncorrelated network ($p = 0$) is $C \simeq 2.9 \cdot 10^{-6}$. Figure 3.9 (b), corresponding to networks of type $\mathfrak{b}$, shows $\mathfrak{u}_m^c$ versus $m+1$ for the values $p = 0.99$ ($C = 0.00040 \pm 0.00005$), $p = 0.9999$ ($C = 0.0522 \pm 0.0007$), and $p = 0.99999$ ($C = 0.344 \pm 0.002$). The mean clustering coefficient C of the uncorrelated network in this case is $C \simeq 0.0001$. In both pictures, the points represent the simulation results and the curves the theory. We note excellent agreement between simulations and Eq. 3.39 for all values of p.

As a next step, we calculate how the number of edges linking two given subsets, for example k-subset a and n-subset b, changes every time the rewiring procedure is executed. We call E_{kn}^{ab} the number of edges connecting k-subset a to n-subset b. In this case the analysis of possibilities yields that E_{kn}^{ab} either increases or decreases by unity, by two, or remains unchanged. After calculating the probabilities of change $E_{kn}^{ab} \to E_{kn}^{ab} \pm 1$ and $E_{kn}^{ab} \to E_{kn}^{ab} \pm 2$, we can obtain the condition in which the probability $\mathfrak{e}_{kn}^{ab} := \langle E_{kn}^{ab}/L \rangle$ that a randomly selected edge of the network connects k-subset a to n-subset b applies in the stationary state. This condition is

$$2p\mathfrak{e}_{kn}^{ab}\left(1 - \textstyle\sum_d \mathfrak{u}_q^d\right) \; + \; (1-p)\left[\left(\mathfrak{e}_{kn}^{ab}\right)^2 + 2\;\mathfrak{e}_{kn}^{ab}\left(1 - Y_k - Y_n\right)\right] \; = \; -\; p\textstyle\sum_d^{a\neq d\neq b}\; \mathfrak{e}_{kq}^{ad}\mathfrak{e}_{nq}^{bd} \; + \tag{3.40}$$

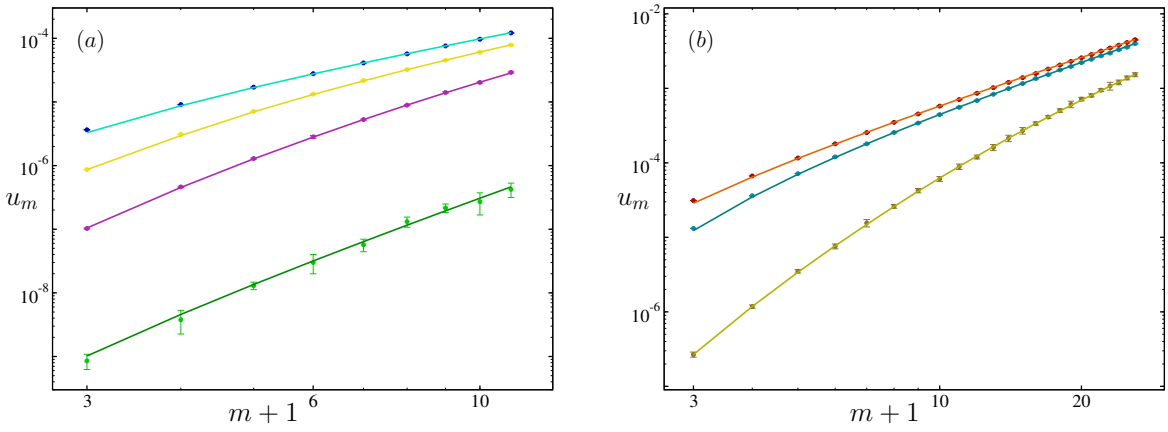

Figure 3.9: $\mathfrak{u}^c_m$ *versus* $m+1$. Panel (a), corresponding to networks type $\mathfrak{a}$: $p = 0.9$, $C \simeq 8.3 \cdot 10^{-6}$, green; $p = 0.999$, $C \simeq 7.9 \cdot 10^{-5}$, violet; $p = 0.9999$, $C \simeq 0.00413$, yellow; $p = 0.99999$, $C \simeq 0.067$, turquoise. Panel (b), corresponding to networks type $\mathfrak{b}$: $p = 0.99$, $C \simeq 0.0004$, dark yellow; $p = 0.9999$, $C \simeq 0.0522$, blue; $p = 0.99999$, $C = 0.344$, red. In both pictures the points correspond to the results of simulations and the curves to the theory.

$$+\, p\left(2Y_k - 2\mathfrak{u}^a_k - \mathfrak{e}^{ab}_{kn}\right)\left(2Y_n - 2\mathfrak{u}^b_n - \mathfrak{e}^{ab}_{kn}\right) \;+\; (1-p)\left[\left(\mathfrak{e}^{ab}_{kn}\right)^2 - 2\mathfrak{e}^{ab}_{kn}\left(Y_i + Y_j\right) \;+\; 4Y_iY_j\right] \quad .$$

Here, sum $\sum_d \mathfrak{u}^d_q$ includes all subsets of the network, while sum $\sum_d \mathfrak{e}^{ad}_{kq}\mathfrak{e}^{bd}_{nq}$ covers all subsets of the network except for subsets a and b (q depends obviously on d). Y_l is give by the expression $Y_l = l(l+1)/2L$. Eq. 3.40 has been calculated assuming that the restrictions for avoiding the appearance of loops and multiedges can be obviated. Unlike the preceding case, the effects of loops and multiedges can however be neglected here, since the probability that multiedges connecting different subsets exist tends to zero in the limit of large networks (loops between different subsets are impossible).

Eq. 3.40 indicates that the solution for $\mathfrak{e}^{ab}_{kn}$ depends on all other connection probabilities $\mathfrak{e}^{uv}_{rs}$ of the network. Thus, we must deal in general with a complex system of nonlinear coupled equations. The problem becomes far easier to solve in the limit of very large networks. In this case, the term $\sum_d \mathfrak{e}^{ad}_{kq}\mathfrak{e}^{bd}_{nq}$ of the equations may be discarded since it yields

$$\left[2Y_k - 2\mathfrak{u}^a_k - \mathfrak{e}^{ab}_{kn}\right]\left[2Y_n - 2\mathfrak{u}^b_n - \mathfrak{e}^{ab}_{kn}\right] - \sum_d^{a\neq d\neq b} \mathfrak{e}^{ad}_{kq}\mathfrak{e}^{bd}_{nq} = \left[2Y_k - 2\mathfrak{u}^a_k - \mathfrak{e}^{ab}_{kn}\right]\left[2Y_n - 2\mathfrak{u}^b_n - \mathfrak{e}^{ab}_{kn}\right] .$$

In order to prove the last equality, let us concentrate on the difference $(2Y_k - 2\mathfrak{u}^a_k) = (1/L)(k(k+1) - 2U^a_k)$. Taking into account that $k(k+1) - 2U^a_k$ is equal to the number of external edges of subset a, we have $k(k+1) - 2U^a_k = \sum_d E^{ad}_{kq}$, where the sum obviously includes all subsets d except for $d = a$. On the other hand, we saw that the TS model works by construction in such a way that the connection probabilities between subsets do not depend on the labels of subsets. Therefore, one can write $k(k+1) - 2U^a_k = \sum_q E^{ad}_{kq} = \sum_q [N_q/(q(q+1))]\, E_{kq} - E_{kk}$. Now, consider the product $(2Y_k - 2\mathfrak{u}^a_k - \mathfrak{e}^{ab}_{kn})(2Y_n - 2\mathfrak{u}^b_n - \mathfrak{e}^{ab}_{kq})$. This product can be rewritten as

$$(2Y_k - 2\mathfrak{u}^a_k - \mathfrak{e}^{ab}_{kn})(2Y_n - 2\mathfrak{u}^b_n - \mathfrak{e}^{ab}_{kn}) = \frac{1}{L^2}\left(\left[k(k+1) \;-\; 2U^a_k\right]\left[n(n+1) \;-\; 2U^b_n\right] + \ldots\right) =$$

$$= \frac{1}{L^2}\left[\left(\sum_q \frac{N_q}{q+1}\, E_{kq}\right) - E_{kk}\right]\left[\left(\sum_p \frac{N_p}{p+1}\, E_{np}\right) - E_{nn}\right] + \ldots =$$

$$= \frac{1}{L^2}\left[\sum_q \frac{N_q}{q+1}\, E_{kq}\right]\left[\sum_p \frac{N_p}{p+1}\, E_{np}\right] + \ldots + \ldots = \frac{1}{L^2}\ \sum_l \left[\frac{N_l}{l+1}\right]^2 E_{kl}E_{nl} + \frac{f(k,n)}{L^2} \quad ,$$

where $f(k,n)$ is a function which groups together the rest of the terms of the product together. Next, we take into consideration the fraction

$$1 \ \geq\ \frac{\left(2Y_k - 2\mathfrak{u}_k^a - \mathfrak{e}_{kn}^{ab}\right)\left(2Y_n - 2\mathfrak{u}_n^b - \mathfrak{e}_{kn}^{ab}\right) \ - \ \sum_d^{a\neq d\neq b} \mathfrak{e}_{kq}^{ad}\mathfrak{e}_{nq}^{bd}}{\left(2Y_k - 2\mathfrak{u}_k^a - \mathfrak{e}_{kn}^{ab}\right)\left(2Y_n - 2\mathfrak{u}_n^b - \mathfrak{e}_{kn}^{ab}\right)} \ \geq$$

$$\geq\ \frac{\sum_l \left(\left[\frac{N_l}{l+1}\right]^2 - 1\right) E_{kl}E_{nl} + f(k,n)}{\sum_l \left[\frac{N_l}{l+1}\right]^2 E_{kl}E_{nl} + f(k,n)} \ =\ \frac{\sum_l \left(\left[\frac{P(l)}{l+1}\right]^2 - \frac{1}{N^2}\right) E_{kl}E_{nl} + \frac{f(k,n)}{N^2}}{\sum_l \left[\frac{P(l)}{l+1}\right]^2 E_{kl}E_{nl} + \frac{f(k,n)}{N^2}} \ .$$

Assuming a stationary degree distribution, this fraction tends to 1 when $N \to \infty$, which finally proves the proposition.

This result ensures that the term $\sum_d^{a\neq d\neq b} \mathfrak{e}_{kq}^{ad}\mathfrak{e}_{nq}^{bd}$ in Eq. 3.40 can be neglected in the limit of large networks. (In fact, the term $\sum_d^{a\neq d\neq b} \mathfrak{e}_{kq}^{ad}\mathfrak{e}_{nq}^{bd}$ can always be discarded provided that the network is not too small, since, as we see from the preceding calculation, the effect of this term in the equations tend to zero as $1/N^2$.) Thus, Eq. 3.40 reduces to the following equation

$$2p\mathfrak{e}_{kn}^{ab}\left[1 - \sum_d \mathfrak{u}_q^d\right] \ + \ (1-p)\left[\left(\mathfrak{e}_{kn}^{ab}\right)^2 + 2\ \mathfrak{e}_{kn}^{ab}\left(1 - Y_k - Y_n\right)\right] \ = \tag{3.41}$$

$$= \ p\left(2Y_k - 2\mathfrak{u}_k^a - \mathfrak{e}_{kn}^{ab}\right)\left(2Y_n - 2\mathfrak{u}_n^b - \mathfrak{e}_{kn}^{ab}\right) \ + \ (1-p)\left[\left(\mathfrak{e}_{kn}^{ab}\right)^2 - 2\mathfrak{e}_{kn}^{ab}\left(Y_i + Y_j\right) \ + \ 4Y_iY_j\right] .$$

when $N \to \infty$. Resolving the equation, we can obtain the probability that a randomly selected edge links two given subsets of the network, one with $k+1$ vertices and the other with $n+1$ vertices. Defining $\hat{L} = L - \sum_d \mathfrak{u}_q^d$ (that is, $\hat{L}$ = total number of external edges in the network), and G_{kn} as

$$G_{kn} \ = \ 1 \ - \ p\left(1 - \frac{\hat{L}}{L}\right) \ + \ p\left(Y_k - \mathfrak{u}_k^a + Y_n - \mathfrak{u}_n^b\right) \tag{3.42}$$

we can write the solution for $\mathfrak{e}_{kn}^{ab}$ as follows

$$\mathfrak{e}_{kn}^{ab} = \begin{cases} \dfrac{4\left(1-p\right)Y_kY_n \ + \ p(2Y_k - 2\mathfrak{u}_k)(2Y_n - 2\mathfrak{u}_n)}{G_{kn} + \sqrt{(G_{kn})^2 - 4p(1-p)Y_kY_n - p^2(2Y_k - 2\mathfrak{u}_k)(2Y_n - 2\mathfrak{u}_n)}} & \text{if} \quad 0 \leq p < 1 \\ 0 & \text{if} \quad p = 1 \ , \end{cases} \tag{3.43}$$

where $\mathfrak{u}_k^c$ is the expected fraction of internal edges of any k-subset in the stationary state. Using Eq. 3.39 to calculate $\mathfrak{u}_m^c(p)$, and implementing it in Eq. 3.43, we can obtain the expected probability $\mathfrak{e}_{kn}^{ab}$.

The degree-degree correlations of the model can now be calculated. For $i \neq j$, the expected connection probabilities $\mathcal{E}_{ij}^t$ are given by $\sum_{a,b} \mathfrak{e}_{ij}^{ab}$, where the sum includes all labels a and b

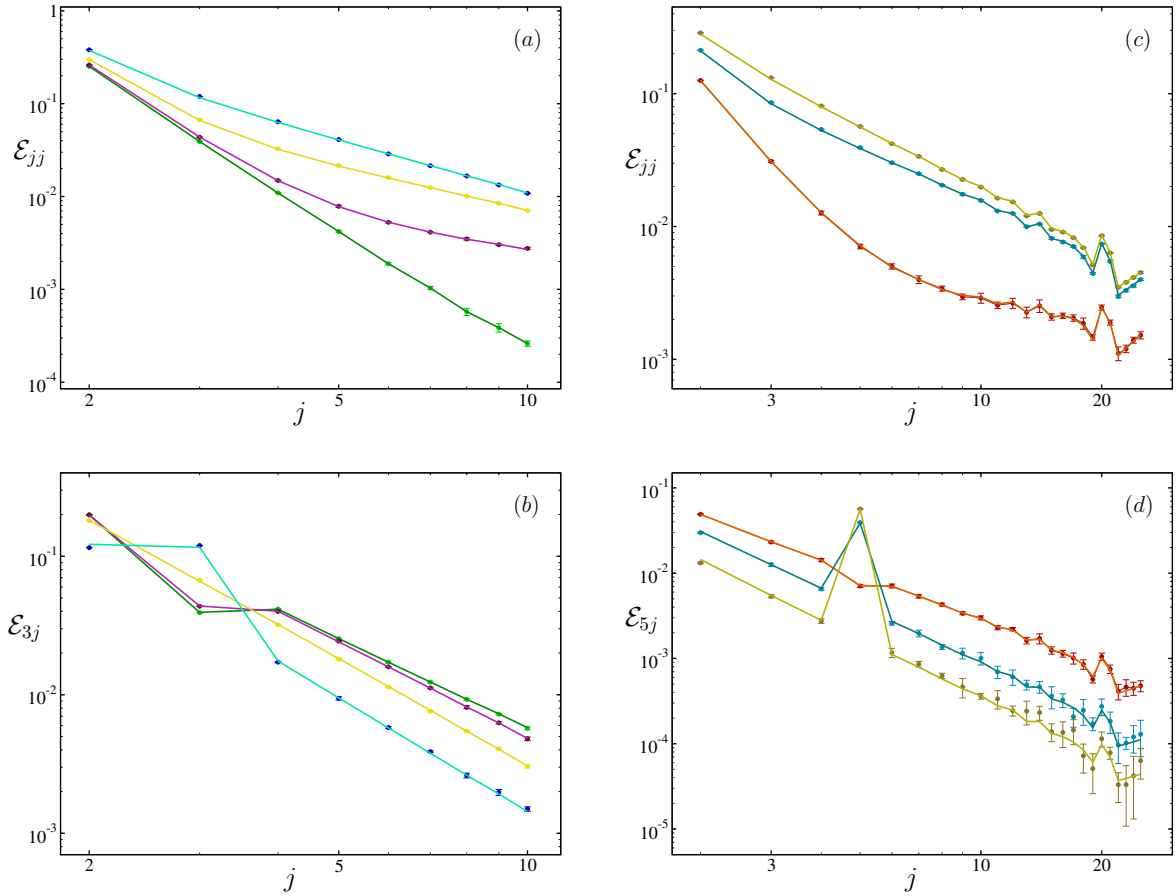

Figure 3.10: Degree-degree correlations of the TS model. Panels (a) and (b), corresponding to networks type $\mathfrak{a}$: $p = 0.9$, $C \simeq 8.3 \cdot 10^{-6}$, green; $p = 0.999$, $C \simeq 7.9 \cdot 10^{-5}$, violet; $p = 0.9999$, $C \simeq 0.00413$, yellow; $p = 0.99999$, $C \simeq 0.067$, turquoise. Panels (c) and (d), corresponding to networks type $\mathfrak{b}$: $p = 0.99$, $C \simeq 0.0004$, dark yellow; $p = 0.9999$, $C \simeq 0.0522$, blue; $p = 0.99999$, $C = 0.344$, red. In all pictures the points correspond to the results of simulations and the curves to the theory.

corresponding to i- and j-subsets of the network. For $i = j$, the expected connection probabilities $\mathcal{E}^t_{ii}$ can be calculated as $\sum_{a,b} \mathfrak{e}^{ab}_{ij}$, for all labels $a \neq b$ corresponding to i-subsets, plus $\sum_c \mathfrak{u}^c_{\mathfrak{i}}$, where the sum covers all labels corresponding to i-subsets. Taking into account that the number of k-subsets is $N_k/(k+1)$, and that $\mathfrak{e}^{ab}_{ij}$ and $\sum_c \mathfrak{u}^c_i$ do not depend on the assigned labels (i. e., we may write $\mathfrak{e}^{ab}_{ij} = \mathfrak{e}_{ij}$ and $\sum_c \mathfrak{u}^c_{\mathfrak{i}} = \sum_c \mathfrak{u}_i$), we obtain the following connection probabilities

$$\mathcal{E}^t_{ij} = \begin{cases} \dfrac{N_i}{(i+1)} \dfrac{N_j}{(j+1)}\, \mathfrak{e}_{\mathfrak{ij}} & \text{if} \quad i \neq j \\ \dfrac{1}{2} \dfrac{N_i}{(i+1)} \left(\dfrac{N_i}{(i+1)} - 1 \right) \mathfrak{e}_{\mathfrak{ij}} + \dfrac{N_i}{(i+1)}\, \mathfrak{u}_i & \text{if} \quad i = j \ , \end{cases} \tag{3.44}$$

expressed as a function of the degree sequence of the network. Eq. 3.44 reduces exactly to Eq. 3.1 when $p = 0$, and takes the value $\mathcal{E}_{ij} = \delta_{ij}$ when $p = 1$. In order to prove the validity of this result, Eq. 3.44, we check simulating the degree-degree correlations of the above considered networks (figure 3.10). The used values of p are particularized in the caption of the picture.

Again, the points represent again the simulation results and the curves the theory. The excellent agreement existing between simulations and theory attest the validity of Eq. 3.41.

To close the study of this type of networks we want to examine them from a different point of view, namely, when considering that TS networks are graphs of ¨supervertices¨. Here, we refer as supervertices to the k-subsets which we have already introduced. Therefore, we simply propose a new level of description of the system. The advantage of describing TS networks as graphs of supervertices is that important properties of the system can be more easily pointed out in this way. For example, we will see that the system suffers drastic changes in some of its structural properties as p increases. Thus, when $p \simeq 0$ the system is a graph of supervertices (more precisely, a *pseudograph* of supervertices) where the fractions of loops and multiedges are negligible (in fact, both fractions vanish completely as $N \to \infty$); in this network of supervertices loops are the internal edges in the different subsets and multiedges are the edges of multiplitiy larger than one sharing the same pair of supervertices. However, when $p \to 1$ the properties of the system are clearly different. In this case, an important part of edges (concretely, $L - \hat{L}$ edges) are loops, i. e, they form part of the ¨internal structure¨ of supervertices, and the rest of edges ($\hat{L}$) interconnect the supervertices. In the extreme case that $p = 1$, $\hat{L} = 0$ and the system becomes an independent set of supervertices-cliques.

Interestingly, when $p \to 1$, the wiring structure of connections between supervertices is degree-degree uncorrelated. Let us discuss this assertion. In order to do this, we consider again the probability that a randomly selected edge of the network connects two given subsets (Eq. 3.43). Now, using the condition $p \to 1$ in Eq. 3.43, and taking into account that $\hat{L} = \sum_m (N_m/(m+1))(2Y_m - 2\mathfrak{u}_m) \lggg 2Y_n - 2\mathfrak{u}_n$ for all n, Eq. 3.43 reduces to $\mathfrak{e}^{ab}_{kn} \simeq (2Y_k - 2\mathfrak{u}_k)(2Y_n - 2\mathfrak{u}_n)/2\hat{L}$, where, as we remember, $(2Y_k - 2\mathfrak{u}_k)L$ and $(2Y_n - 2\mathfrak{u}_n)L$ are no more than the degrees of the supervertices. Taken together, when $p \to 1$, the number of edges which, on average, interconnect two given supervertices, is $1/2$ times the product of their expected degrees over $\hat{L}$. Furthermore, numerical simulations indicate that the degree of any supervertex is always equal or very close to its expected degree. Thus, denoting by $\hat{N}_d$ the number of supervertices of degree d, the connection probabilities between supervertices will be given by

$$\hat{\mathcal{E}}_{qd} = \begin{cases} \dfrac{1}{\hat{L}} \hat{N}_q \hat{N}_d \dfrac{qd}{2\hat{L}} & \text{if} \quad q \neq d \\ \dfrac{1}{2\hat{L}} \hat{N}_q \hat{N}_d \dfrac{qd}{2\hat{L}} & \text{if} \quad q = d \end{cases} \qquad \longrightarrow \qquad \hat{\mathcal{E}}_{qd} = (2 - \delta_{qd}) \frac{q\hat{P}(q) d\hat{P}(d)}{\langle q \rangle \langle d \rangle} \ , \tag{3.45}$$

where $\langle d \rangle$ is the first moment of the degree distribution $\hat{P}(d)$ of supervertices. We indicate with the "hat" in $\hat{\mathcal{E}}_{qd}$ that this result corresponds to the pseudograph of supervertices. Thus, the network of supervertices is practically uncorrelated when $p \to 1$. We check the goodnes of Eq. 3.45 in figure 3.11. We compare in the pictures the measured connection probabilities $\hat{\mathcal{E}}_{qd}$ between supervertices with the corresponding theoretical prediction given by Eq. 3.45; in order to obtain the theoretical probabilities, we measure the supervertex degree distribution of the network under consideration and then we use Eq. 3.45. All results correspond to the above networks of type $\mathfrak{a}$ and $\mathfrak{b}$, and are averaged over five realizations of the process. In panels (a)-(b) and (c)-(d) we respectively plot $\hat{\mathcal{E}}_{dd}$ (network $\mathfrak{a}$, $p = 0.9999$), $\hat{\mathcal{E}}_{8d}$ (network $\mathfrak{a}$, $p = 0.99999$), $\hat{\mathcal{E}}_{dd}$ (network $\mathfrak{b}$, $p = 0.9999$), and $\hat{\mathcal{E}}_{6d}$ (network $\mathfrak{b}$, $p = 0.99$) as a function of the supervertex degree d. The points of the pictures correspond again to the simulation results, while the curves correspond to the connection probabilities that the networks should have if they were uncorrelated. We note excellent agreement. Panels (c) and (f) show the behavior the nearest neighbors' average degree function $\bar{k}_{nn}(d)$ of, respectively, networks type $\mathfrak{a}$ and $\mathfrak{b}$. Here, points which the corresponding error bars are calculated using Eq. 3.11 from the measured $\hat{\mathcal{E}}_{qd}$, while linies are obtained using

the same equation but from the connection probabilities given by Eq. 3.45. (Points without error bars correspond to the measured $\bar{k}_{nn}(d)$ too, but they are not statistically significant since they correspond to measures which can be found in only one network from the five ones which we use to obtain the average.) The fact that $\bar{k}_{nn}(d)$ is a constant function in all considered cases demonstrates the uncorrelatedness of the connections between supervertices. On the other hand, we want to remark that the values of p for which it can already be said that the system is uncorrelated (taking into account Eq. 3.45, from $p \geq 0.99$ on, for instance) correspond to relative small values of transitivity degree of the network. In fact, real networks typically show values of C larger than the ones from which on the nework of supervertices is uncorrelated. Note also that, in order to build networks with a given C but different order, we need constantly larger values of p as N increases, which ensures that large TS networks having non-negligible clustering coefficients will always be uncorrelated.

On the other hand, the supervertex degree distribution $\hat{P}(d)$ also changes drastically as p increases. When $p \simeq 0$ it verifies $\hat{P}(d) \sim k(k+1)P(k)/(k+1) = kP(k)$, for all d such that $d = k(k+1)$; otherwise, $\hat{P} = 0$. In fact, there is usually a very small number of supervertices of degree $k(k+1)-1$ or $k(k+1)-2$, for all k. This property corresponds to the fact that a few supervertices have one or two internal edges even when $p \simeq 0$. The fraction of such supervertices, however, tends to zero when $N \to \infty$. As $p \to 1$ the degree of supervertices decreases progressively until it vanishes at exactly $p = 1$.

Let us analyze how the $\hat{P}(d)$ varies as p changes. To this end, consider the probability $p(d, s, i, t)$ that at time t supervertex s composed of $i+1$ vertices has degree q. From now on, we refer to a supervertex composed of $i+1$ vertices as a supervertex of ¨characteristic¨ i. The key factor for calculating how $p(d, s, i, t)$ changes with time is that the degree of any supervertex in the pseudograph can only increase or decrease by two unities every time the rewiring process is applied. The master equation governing this probability $p(d, s, i, t)$ is:

$$\begin{aligned} p(d,s,i,t+1) \;=\; & p(d,s,i,t)\,[1-\mathcal{P}^s_{+2}(d,i)-\mathcal{P}^s_{-2}(d,i)] + \\ & + p(d-2,s,i,t)\;\mathcal{P}^s_{+2}(d-2,i) \;+\; p(d+2,s,i,t)\;\mathcal{P}^s_{-2}(d+2,i) \end{aligned} \tag{3.46}$$

where $\mathcal{P}^s_{+2}(d,i)$ and $\mathcal{P}^s_{-2}(d,i)$ respectively are the probabilities that a supervertex of degree d and characteristic i undergoes an increase or decrease in its degree. These probabilities are given by the expressions

$$\mathcal{P}^s_{+2}(d,i) = \mathcal{P}^t_{-1}\left(\frac{i(i+1)-d}{2}, i\right) \tag{3.47}$$

and

$$\mathcal{P}^s_{-2}(d,i) = \mathcal{P}^t_{+1}\left(\frac{i(i+1)-d}{2}, i\right) \tag{3.48}$$

where probabilities $\mathcal{P}^s_{+2}(d,i)$ and $\mathcal{P}^s_{-2}(d,i)$ are obviously related to $\mathcal{P}^t_{-1}(d,i)$ and $\mathcal{P}^s_{+1}(d,i)$ because of the relation $i(i+1)-2U = d$ (where U is the number of internal edges of the supervertex). Note that none of either probabilities would be well defined if d were odd; this problem does not exist since the degree d of any supervertex is always even, given that $i(i+1)$ is an even integer $\forall\, i$. The probability that a randomly selected supervertex of characteristic i has degree d when the system reaches the stationary state can be obtained from equation

$$\hat{P}(d,i) = \lim_{t\to\infty} \frac{1}{N_i} \sum_s p(d,s,i,t)\;, \tag{3.49}$$

where the sum evidently includes every supervertex s of characteristic i. If we now evaluate the sum $\sum_s$ in both sides of Eq. 3.46, and use the previous definition of $P(d,i)$, we obtain the

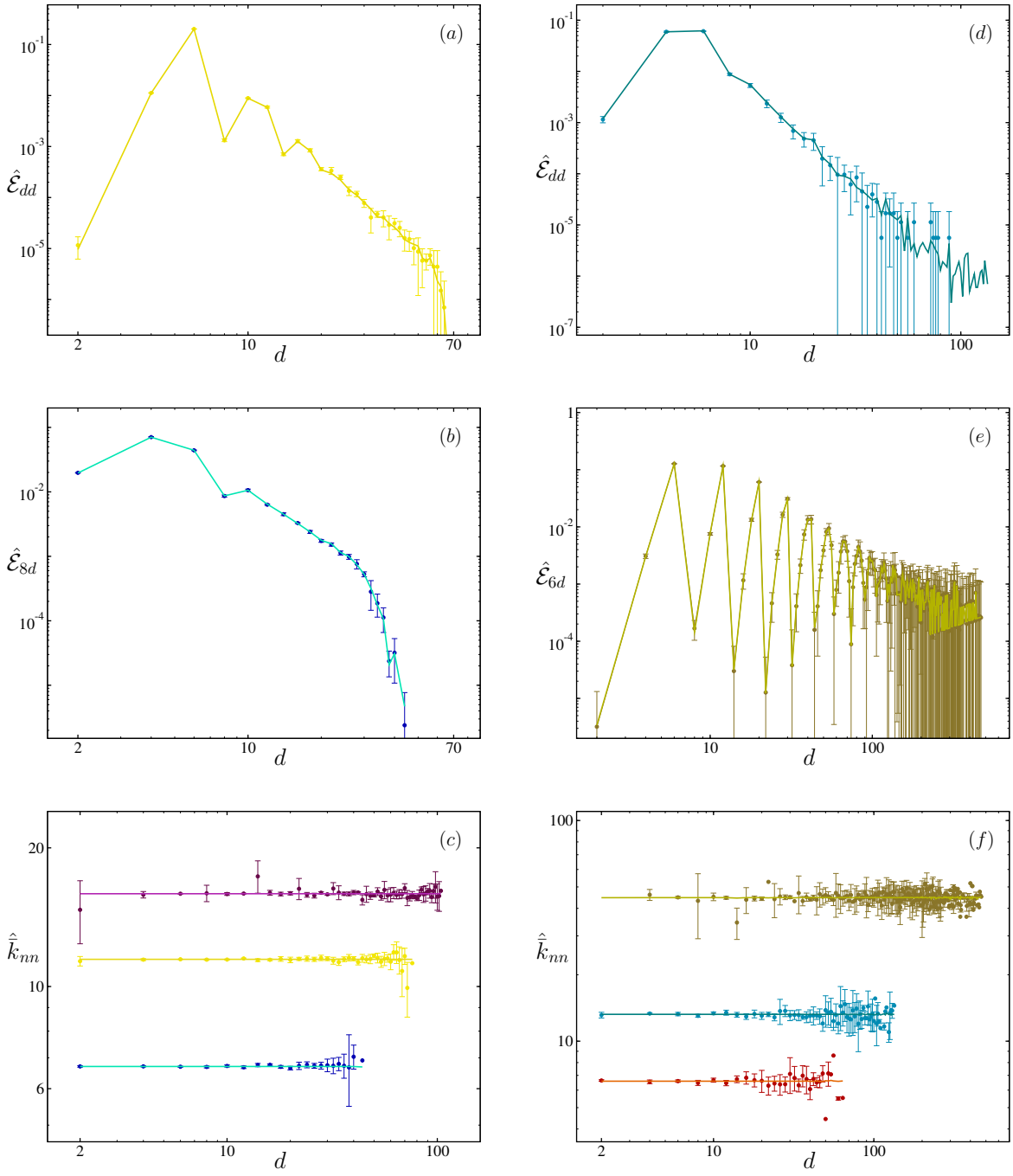

Figure 3.11: Panels (a-b) and (d-e): Different supervertex connection probabilities $\mathcal{E}_{qd}$ as a function of d. Panel (a): network $\mathfrak{a}$, $p = 0.9999$, $C = 0.00413$. Panel (b): network $\mathfrak{a}$, $p = 0.99999$, $C = 0.0677$. Panel (d): network $\mathfrak{b}$, $p = 0.9999$, $C = 0.0522$.Panel (e): network $\mathfrak{b}$, $p = 0.99$, $C = 0.00040$. The lower panels (c) and (f) show how function $\bar{k}_{nn}(d)$ behaves as a function of the supervertex degree d. Panel (c), corresponding to networks of type $\mathfrak{a}$: from top to bottom, $p = 0.999$ ($C = 7.9 \cdot 10^{-5}$), $p = 0.9999$, and $p = 0.99999$. Panel (e), corresponding to networks of type $\mathfrak{b}$: from top to bottom, $p = 0.99$, $p = 0.9999$, and $p = 0.99999$ ($C = 0.344$). Note that the fact that the network of supervertices is a pseudograph yields $\bar{k}_{nn}(d)$ constant even for large degrees d.

following equation linking the probabilities $P(d,i)$, $P(d+2,i)$, and $P(d-2,i)$

$$\hat{P}(d,i)[\mathcal{P}^s_{+2}(d,i)+\mathcal{P}^s_{-2}(d,i)]=\hat{P}(d+2,i)\mathcal{P}^s_{-2}(d+2,i)+\hat{P}(d-2,i)\mathcal{P}^s_{+2}(d-2,i)\ . \tag{3.50}$$

Taking into account that $\mathcal{P}^s_{-2}(0,i)=0$, $\mathcal{P}^s_{+2}(i(i+1),i)=0$, and the normalization condition $\sum_d \hat{P}(d,i)=1$, the last equation can be solved in an iterative way, yielding the following expression for $\hat{P}(d,i)$

$$\hat{P}(d,i)=\frac{\displaystyle\prod_{l=0}^{d/2-1}\mathcal{P}^s_{+2}(2l,i)\prod_{j=d/2+1}^{i(i+1)/2}\mathcal{P}^s_{-2}(2j,i)}{\displaystyle\sum_{k=0}^{i(i+1)/2}\prod_{n=0}^{k-1}\mathcal{P}^s_{+2}(2n,i)\prod_{m=k+1}^{i(i+1)/2}\mathcal{P}^s_{-2}(2m,i)}\ . \tag{3.51}$$

Once the probability $\hat{P}(d,i)$ is known we can obtain the supervertex degree distribution $\hat{P}(d)$ of the pseudograph making use of

$$\hat{P}(d)=\frac{\displaystyle\sum_j \frac{N_j}{j+1}\hat{P}(d,j)}{\displaystyle\sum_l \frac{N_l}{l+1}}=\frac{\displaystyle\sum_j \frac{P(j)}{j+1}\hat{P}(d,j)}{\displaystyle\sum_l \frac{P(l)}{l+1}}\ . \tag{3.52}$$

Thus, the supervertex degree distribution, $\hat{P}(d)$, depends on the degree distribution $P(k)$ of the original network, and on the model parameter p. By varying p in Eq. 3.52, the supervertex degree distribution changes ranging from that corresponding to $p=0$, where almost all supervertices have a degree $d=k(k+1)$, with k integer, to $p=1$, where all supervertices have $d=0$.

In order to verify the validity of the preceding calculation, we compare in figure 3.12 (*a-f*) the theoretical degree distribution $\hat{P}$ given by Eq. 3.52 with the experimental supervertex degree distribution obtained from simulations. The results correspond to some of the previously considered networks (see caption of the picture). We observe two things: i) that the agreement between simulations and theory is of excellent standard, and ii) that for all values of p considered, the supervertex degree distribution decays as a power law whose exponent decreases as p grows, or even more rapidly. For instance, consider figures 3.12 (*d-f*) where are plotted the results corresponding to network $\mathfrak{b}$. We see that for $p=0.99$ the supervertex degree distribution $\hat{P}(d)$ bears and ¨oscillating¨ function which more or less decays following a power-law of exponent $\gamma\simeq -2.4$; however, for $p=0.9999$ and $p=0.99999$, the pictures show almost perfect power-law decayings governed respectively by the exponents $\gamma\simeq -2.75$ and $\gamma\simeq -3.2$. Network $\mathfrak{a}$ presents the same qualitative behavior. In this case, for $p=0.999$ ($C\simeq 0.00008$) the exponent of the power law is $\gamma\simeq -3.15$, for $p=0.9999$ ($C\simeq 0.0041$) is $\gamma\simeq -3.3$, and for $p=0.99999$ ($C\simeq 0.068$) is $\gamma\simeq -3.6$. Thus, we have encountered networks with such a vertex degree distribution $P(k)$ that its second moment diverges ($P(k)\sim k^{-\beta}$ with $\beta\leq 3$), but with such a supervertex degree distribution that its second moment is finite ($\hat{P}(d)\sim d^{-\gamma}$ with $\gamma>3$). This fact will be especially relevant when dealing with percolation properties of these networks (see chapter 4). We notice that this effect even occurs if the mean clustering coefficient is small. The effect is not surprising. Sure enough, consider the solution for $\hat{P}(d,i)$ given by Eq. 3.49. Analyzing this solution, we can observe that when $p\to 1$ as $p\sim 1-1/3L$, $\hat{P}(d,i)$ exhibits an sharp peak around $d\sim i$ so that $\hat{P}(d,i)$ can roughly be approximated by $\hat{P}(d,i)\sim\delta_{di}$. (Note that the probability $\hat{P}(d,i)$ does not depend on the degree distribution $P(k)$, but only on the

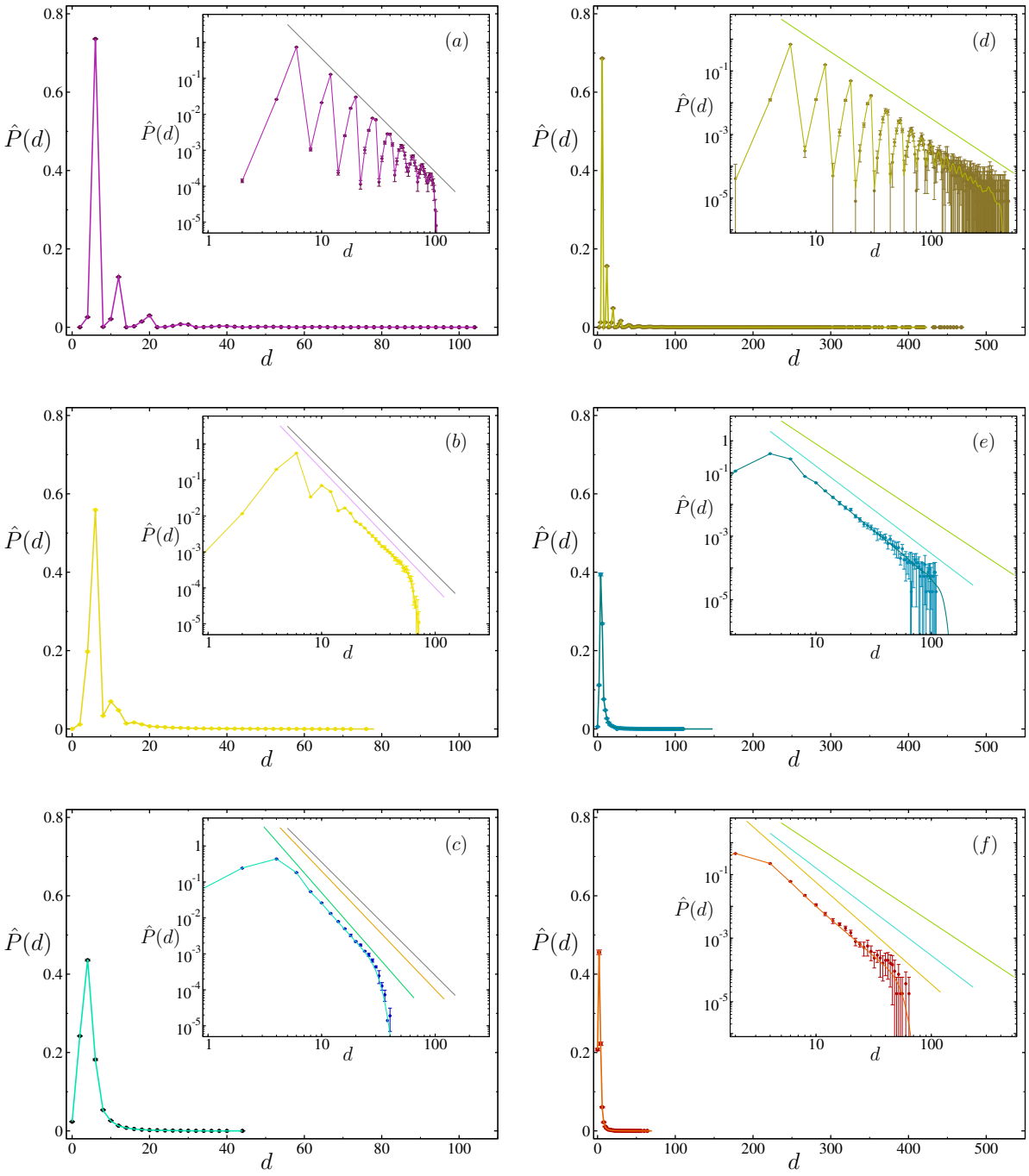

Figure 3.12: Supervertex degree distribution $\hat{P}(d)$ of some of the TS networks already considered: panels $(a\text{-}c)$, networks type $\mathfrak{a}$; panels $(d\text{-}f)$, networks type $\mathfrak{b}$. Panel a: $p = 0.999$; panel b: $p = 0.9999$; panel c: $p = 0.99999$; panel d, $p = 0.99$: panel e: $p = 0.9999$; panel f: $p = 0.99999$. In all pictures the points correspond to the results of simulations and the curves to the theory. We note excellent agreement between simulations and theory. The supervertex degree distribution approximately decays as a power law for all values of p considered (see text for details).

parameter p. Therefore, this result is valid regardless of the network considered). Implementing $\hat{P}(d,i) \sim \delta_{di}$ in Eq. 3.52, we obtain $\hat{P}(d) \propto P(d)/d$. That is, if the original network is scale-free, $P(k) \sim k^{-\beta}$, then the corresponding network of supervertices will also be scale-free $\hat{P}(d) \sim d^{-\gamma}$ and the exponent γ will approximately equal $\gamma \simeq \beta - 1$. This result is consistent with our simulations. Of course, for much more large values of p we expect that the scale-free character of $\hat{P}(d)$ disappears and the degree of supervertices tend rapidly to zero.

To conclude this section, we hint at that the TS algorithm can also be used to generate networks exhibiting nontrivial community structure. One must simply change the labeling procedure so that all vertices which must belong to the same community have the same label. By varying the probability p, one can change the *modularity*[8] of the network. Although the topology of networks thus generated will probably differ from that found in real community networks, they could be very useful for checking the quality of the recent algorithms capable of detecting community structure in networks [Newman, 2004, Fortunato et al., 2004, Newman and Girvan, 2004, Clauset et al., 2004]. The algorithm could also be used in immunology: Systems exist in which individuals are usually grouped into small sets of very strongly connected persons. These systems are more conveniently modeled as networks of supervertices. With respect to this, we notice that, selecting appropriately the degree sequence of the original network and the assignment of cliques, one can construct a multigraph composed of supervertices (cliques) obtaining the supervertex degree sequence one desires.

3.4 Topological properties

The principal goal of this chapter is to investigate how vertex-pair correlations change the topological properties of networks. For this purpose, we introduce in the preceding sections three algorithms capable of restructuring the connections between vertices in different ways. Here, we analyze the structure of the networks constructed by applying these algorithms to scale-free constructions. We focus the study on the average path length, transitivity, and tomography of shell structure around an arbitrary vertex. The results demonstrate that vertex-pair correlations strongly influence the topological properties of networks.

Average path length. The average path length of a network is the average distance between every pair of vertices of the network. Uncorrelated scale-free networks (actually, the most networks we know, except for lattices), show a very small path length which typically grows logarithmically with the network order (small-world behavior). This behavior is actually observed in all networks we know, except for networks like lattices. However, the peculiar distribution of connections influences the calculation of distances. Therefore, we expect that a network exhibits different average distances as a function of their mixing. We ask here how the mean distance varies in networks when vertex-pair correlations change. The algorithm we use to measure the average path length of a network is trivial: starting from an initial vertex (labeled 0), we pass to all vertices connected to it (vertices of the first generation –neighbors– labeled 1), then to vertices of the second generation (labeled 2), etc, until all vertices are labeled. The mean distance between this vertex (labeled 0) and any other given vertex of the network is then the sum of all values of these labels divided by $N-1$. This procedure is repeated for each vertex, and the overall mean value, which corresponds to the searched average path length l, is evaluated. Since

[8]Modularity is one of the quantities introduced in corresponding studies to measure the degree of clustering in community networks. It basically measures the fraction of edges in the network that connect different *modules* or communities to each other; the smaller the fraction of connecting edges the more pronounced the community structure is.

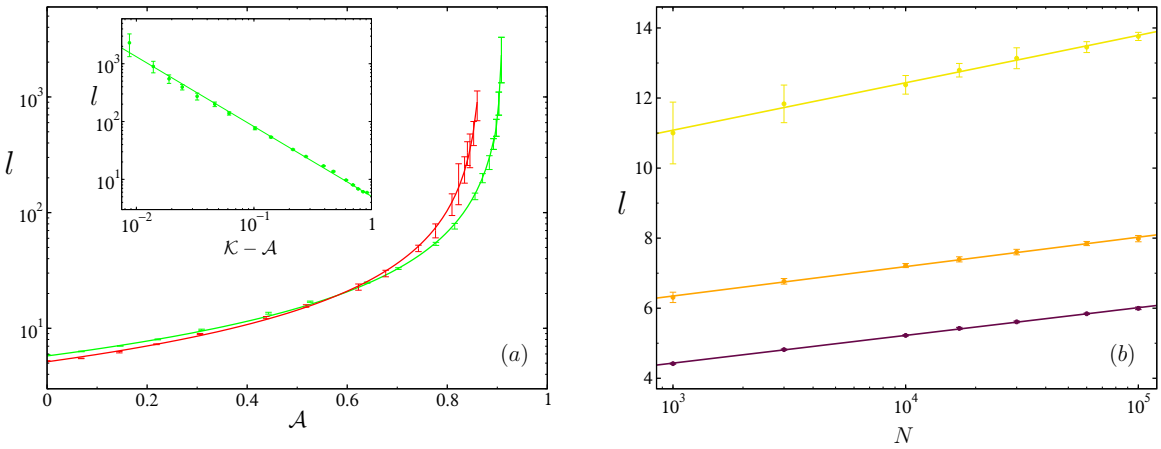

Figure 3.13: AS model. (a): Average path length l *versus* the assortativity coefficient $\mathcal{A}$ for two AS-Barabási-Albert networks, one of 10^5 vertices (green) and one another of 10^4 (red). We note that l grows rapidly when $\mathcal{A}$ increases. The average path length is plotted on double logarithmic scales as function of $\mathcal{K} - \mathcal{A}$ for the larger network ($N = 10^5$) in the inset. Here is $\mathcal{K} = 0.917$. The slope of the straight line is -1.2. (b): Average path length l as function of network order, N, for three different assortativated AS networks: From bottom to top: $\mathcal{A} = 0$, $\mathcal{A} = 0.221$, and $\mathcal{A} = 0.443$. Note the logarithmic scale.

the average distance is only defined for fully connected networks, we will refer to the average path length of the largest component of the network when we speak about the average distance of a disconnected network.

Let us begin with the AS model. We apply this algorithm to Barabási-Albert scale-free networks of different orders. When stated without any specification, the networks agree to $L = 2 \cdot N$ edges ($m = 2$), where N is the order of the network considered. We denote by AS-Barabási-Albert networks the networks constructed using the Barabási-Albert prescription to which the AS algorithm has been applied. Numerical simulations show that l grows rapidly when the assortativity of the network increases (figure 3.13, panel a); it increases until it reaches values hundreds times larger than the ones corresponding to uncorrelated networks when the coefficient of assortativity, $\mathcal{A}$, approaches its maximal value. The plotted curves correspond to simulation results based on two AS-Barabási-Albert networks, one of 10^4 vertices and another of 10^5 vertices (see caption). A detailed analysis of curves shows that both curves can be successfully fitted using the function $l \propto (\mathcal{K} - \mathcal{A})^{-\gamma}$ with $\mathcal{K} = 0.864$ (for the small network, $N = 10^4$) and $\mathcal{K} = 0.917$ (for the large network, $N = 10^5$) -corresponding to the maximal values of $\mathcal{A}$ attainable in the networks- and with $\gamma = 1.2$. In the inset of the figure we re-plot one of the curves ($N = 10^5$) using more convenient scales on the axis; the inset confirms the fit.

This behavior of the average path length seems to be characteristic of networks where vertices are grouped forming "communities". Sure enough, as the clustering grows, the average distance increases [Gómez-Gardeñes and Moreno, 2004]. The explanation is simple: As the clustering increases, the vertices belonging to a given community tend to connect more and more between themselves, and, consequently, the number of paths connecting the different communities decreases. Thus, the underlying structure compels to travel along many edges to go from vertices of one community to vertices of another, which leads to an increase of the mean distance (l) in community networks. In assortative networks, vertices of similar degree tend to group together forming, as a result of this, communities composed of similar degree vertices. The community

structure is even more peculiar in this case than in normal community networks: communities of vertices of degree k communicate preferably to other communities composed of vertices of similar degree ($\simeq k$), but hardly to communities of vertices having degrees either much larger or much smaller than k. Consequently, the geodesics jointing high-degree vertices to low-degree vertices are typically very long in strongly assortative networks, since the connecting paths usually pass along many communities in-between. Near the extreme case $p = 1$ (maximal assortativity), such paths must inevitably pass through almost all communities in-between, which results in an enormous value of the mean path length, l.

Assortative networks present large average path lengths. We ask next whether they are still small worlds or not. Interestingly, AS-scale-free networks are certainly small worlds, i.e., they exhibit the typical logarithmic dependence of l on the network's order N. Figure 3.13 (b) shows this behavior for three AS-Barabási-Albert networks, each one with different value of assortativity. The error bars result from averaging over ten independent realizations of the algorithm. The simulation results demonstrate their small world character. The small-world behavior is preserved for all tested values of $\mathcal{A} \leq 0.6$ (value which is considerably larger than those observed in real assortative networks). Thus, assortative networks could be considered as "large" small worlds.

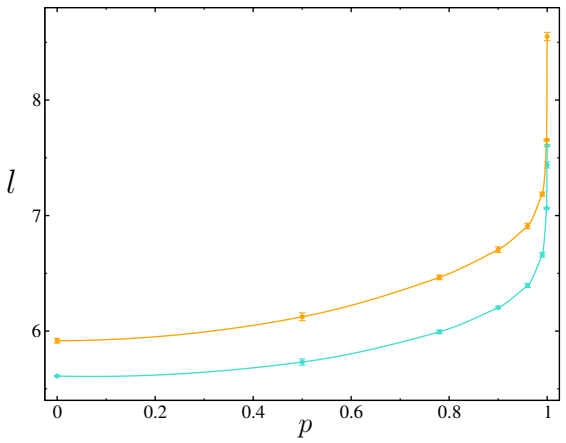

Figure 3.14: DS model. Shortest average path length *versus* the parameter p, which we use to characterize the degree of dissortativity of the network. The upper curve corresponds to a DS-Barabási-Albert network of $N = 10^5$ vertices, while the lower curve to one of $N = 3 \cdot 10^4$.

In relation to various real world systems, like the different co-authorship networks (physics, biology, mathematics, etc.), the film actor collaboration network, etc (all of them assortative networks), we observe somewhat smaller average path lengths than the ones found here for AS networks [Newman, 2003a]. We attribute this finding to the fact that the mean degree of such networks is 2 to 4 times larger than in our simulated case ($\langle k \rangle = 4$). Therefore, one has to be cautious about comparing absolute numerical values.

Consider now the DS model. Figure 3.14 (b) shows the behavior of l when networks are dissortatively mixed. Here, we apply the DS algorithm to two scale-free Barabási-Albert networks, one of $3 \cdot 10^4$ vertices and another of 10^5; the number of edges is $L = 2 \cdot N$ ($m = 2$), as before. In the picture we plot the average shortest path length as function of the parameter p. (Although the coefficient $\mathfrak{D}$ would probably be a more suitable parameter to measure the degree of dissortativity of a network, we discard its use here since $\mathfrak{D}$ has not been introduced in the literature yet.) The picture shows that the average distance l grows gradually when the dissortativity of the network increases. Unlike the assortative case, the increase of l is moderate, and its maximal value is not much higher than in an uncorrelated network.

At last, we discuss how the average distance l varies as function of C, the mean clustering coefficient, in highly triangulated (TS) networks. In this case, contrarily to what occurs in the

assortative and dissortative cases, where the network remains fully connected even when the AS and DS algorithms are applied with very large values of p, the network here tends to break into isolated components when C increases sufficiently. Figure 3.15 (a) and (b) respectively show how the fraction of vertices M, and the average path length l of the giant component change when the mean clustering coefficient grows. All simulation results are also averaged over ten independent realizations of the algorithm, and are based on Barabási-Albert constructions of $3 \cdot 10^4$ vertices (and $6 \cdot 10^4$ edges, $m = 2$). As we already saw, the networks remain connected until the mean clustering coefficient reaches the value $C \simeq 0.12$; from this value on the network tends to separate in disconnected clusters and the mass of the giant decreases. The average path length, l, grows monotonously when C increases. It increases until $C \sim 0.9$ (a very large value in comparison to real networks), despite the continuous reduction of the order of the giant component; for larger C, the mean distance l rapidly falls (last point in figure 3.15).

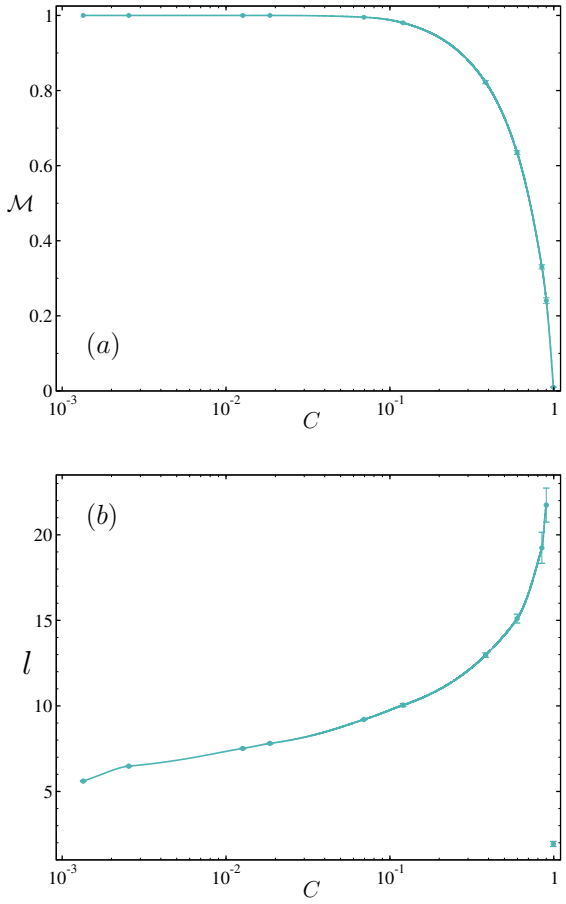

Figure 3.15: TS model. (a): Fraction of vertices in the giant component, $\mathcal{M}$, as function of the mean clustering coefficient C. Note that the network remains connected until $C \leq 0.12$ (two orders of magnitude larger than the value corresponding to the uncorrelated case $C = 0.0013$). (b): Average shortest path length l *versus* the mean clustering coefficient C. There is a sharp increase of the average distance when C increases until $C \sim 0.9$ despite the reduction of the giant component. For larger degrees of clustering the mean distance l falls rapidly.

To summarize, we show that correlations modify the average path length of networks considerably. This change, however, does not affect their small world behavior (except for, perhaps, extreme assortative networks). It is also quite interesting to note that l is, for all correlated networks examined, larger than the corresponding l for uncorrelated networks.

Transitivity. We study here how correlations influence clustering coefficients C and $C(k)$. As above, we begin with the AS model. Figure 3.16 (a) shows the variation of the mean clustering coefficient C as a function of quantity $\mathcal{A}$ for two AS-Barabási-Albert networks, one of 10^5 vertices (upper curve) and another of 10^6 vertices (lower curve); in both cases $L = 2 \cdot N$

($m = 2$). The results are averaged here over five (largest network, $N = 10^6$) and ten (smallest network, $N = 10^5$) independent realizations of the algorithm. The curves indicate that the mean clustering coefficient increases when the assortativity grows. However, in spite of showing an increase by more than one order of magnitude, C is still much smaller than the typical values found in real networks ($C \geq 0.1$). The latter is, however, not surprising, since real networks have a much more intricate structure than our simple and purely mathematical model: they are very probably governed by the metrics of the underlying space, as well as by peculiar characteristics of vertex connections.

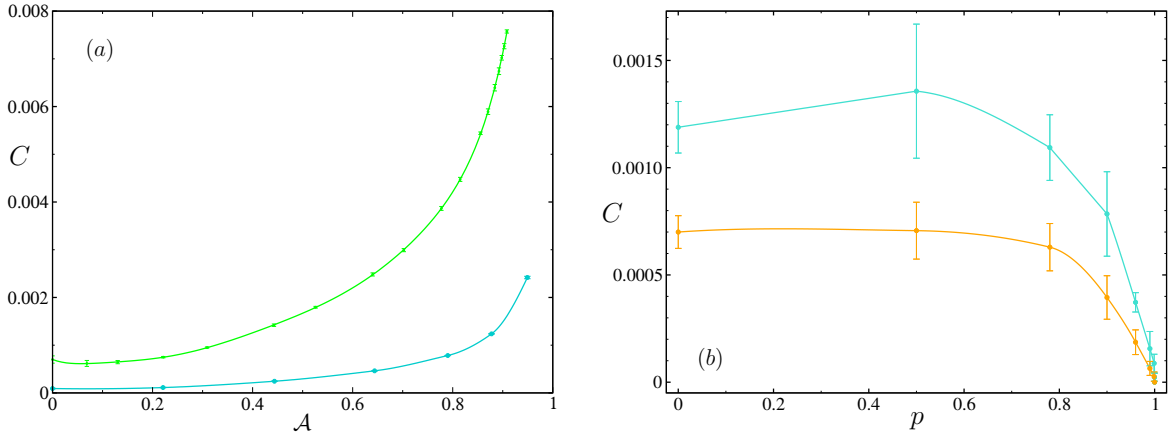

Figure 3.16: (a): AS model. Mean clustering coefficient C *versus* the coefficient of assortativity $\mathcal{A}$. The upper curve corresponds to a network with $N = 10^5$ vertices and the lower to another with $N = 10^6$ vertices. (b): DS model. C as function of the parameter p (where we use p as a measure of the dissortative mixing in the network). $N = 10^5$ (orange), $N = 3{\cdot}10^4$ (blue). Both curves vanish exactly at $p = 1$ (see text for details).

On the contrary, in DS networks the mean clustering coefficient decreases as p grows. Again, we use p in figure 3.16 (b) to plot the mean clustering coefficient as a function of the degree of dissortativity. The networks studied are based on Barabási-Albert constructions with $N = 3{\cdot}10^4$ and $N = 10^5$. The simulations indicate that the degree of transitivity does not vary considerably when the dissortativity increases from $p = 0$ (uncorrelated networks) to moderate values of p (not too strongly dissortative networks). When the dissortativity increases further, C begins to decrease and eventually vanishes. In effect, a detailed counting of triangles indicates that totally dissortative DS networks ($p = 1$) have no circuits of length three. The explanation is not difficult: A vertex may form a part of a triangle if at least an edge connecting two of its neighbors exists. Now, consider a low-degree vertex in a maximally dissortative network. All its neighbors will be vertices of large degree, which cannot be connected to each other as high degree vertices connect only to low-degree vertices. Consider a vertex of large degree. It will connect only to low-degree vertices, which cannot be connected between themselves because all vertices of small degree connect to high degree vertices. Note that cycles of length four may however exist even in ample number. Triangles may then be formed only by moderate degree vertices; nevertheless, the probability is small. For $p < 1$, the situation does not change dramatically. Simulations corresponding to strongly but not maximally dissortative DS networks show the existence of a few triangles composed exclusively of vertices of degrees 4, 5, or 6 (i. e., the moderate degree vertices for which a certain degree of assortativity exists), but no triangles containing low or high degree vertices.

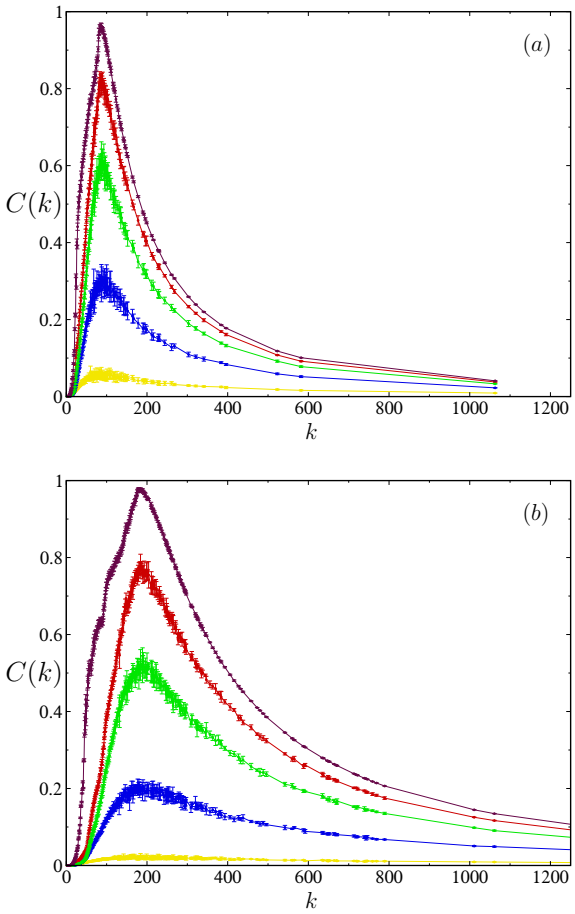

Figure 3.17: AS model. Degree-dependent clustering coefficient $C(k)$ as function of the degree of vertices k. The pictures correspond to AS-Babarási-Albert networks with $N = 10^5$ (panel a) and $N = 10^6$ (panel b). Each curve corresponds to a different degree of assortativity: $\mathcal{A} = 0$, $\mathcal{A} = 0.221$, $\mathcal{A} = 0.443$, $\mathcal{A} = 0.640$, $\mathcal{A} = 0.777$, and $\mathcal{A} =$ maximum degree ($\mathcal{A} = 0.917$ in panel a, $\mathcal{A} = 0.949$ in panel b), from bottom to top. (The curve corresponding to $\mathcal{A} = 0$ (uncorrelated case) is not visible, since this curve and the inferior frame of the pictures coincide.)

The analysis of the degree-dependent clustering coefficient $C(k)$ indicates that in DS networks $C(k) \to 0$, for all k, as the dissortativity increases. The only remarkable characteristic is the already commented fact that a small peak appears at $k \simeq 5$ when p is close but not equal to unity. The study of the local coefficient $C(k)$ for the AS model offers, however, very interesting features. Figure 3.17 shows the behavior of $C(k)$ as a function of k for the same AS-Barabási-Albert networks we considered above when studying C (figure 3.16 a). The simulation results show the appearance of a peak (around $k = 90$ for the network of 10^5 vertices, panel a, and around $k = 185$ for the network of 10^6 vertices, panel b) whose height increases with the assortativity of the network. The peak (probably a finite size effect) moves to larger k when the order of the network increases. Thus, contrarily to uncorrelated networks (where $C(k)$ does not depend on k [Dorogovtsev, 2004]), assortative AS networks present a strong tendency to "triangulate" for some values of k.

A careful counting of triangles indicates in this case that $C(k = 2) = 0$ ($k = 2$ corresponds to the minimum degree of the networks considered) when $A \simeq 1$. This is not surprising since in strongly assortative networks almost all vertices of degree $k = 2$ are connected between themselves, forming one or several circuits of length usually larger than three. Bear in mind that in AS networks vertices of same degree connect to each other at random. Thus, the probability that a triangle composed exclusively of vertices of degree two appears in a completely assortative network is not different to the probability that it emerges in a random network consisting only of N_2 vertices of degree two. This probability is certainly very small when N_2 is large. The same arguments explain the marginal values of $C(k)$ when k is small. Consider now the vertices whose

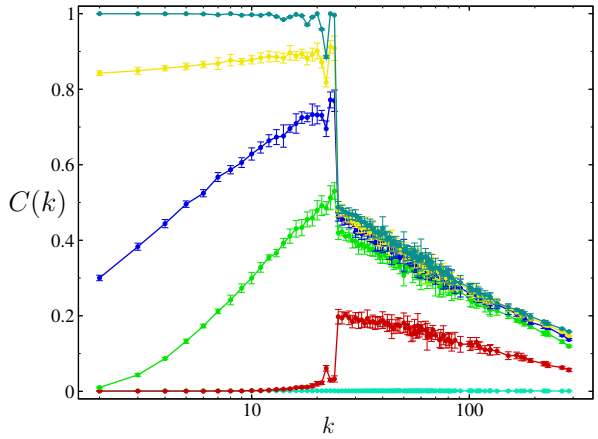

Figure 3.18: TS model. $C(k)$ as function of k for different triangulated TS-Barabási-Albert networks. From bottom to top: $C = 0.00134$ (uncorrelated case), $C = 0.00255$, $C = 0.0695$, $C = 0.383$, $C = 0.846$, and $C = 0.994$ (maximum C). All considered networks have exactly the same degree sequence. The order of the networks is $N = 3 \cdot 10^4$.

degree i satisfies $i \simeq N_i$. For all $k \simeq i$, $C(k)$ increase evidently as the assortativity grows, because cliques composed of vertices of degree $k \simeq i$ tend to be formed. To this respect, note that the larger the network's order N, the larger is the degree i at which the condition $i \simeq N_i$ is fulfilled. This feature explains the movement of the peak to a larger k as $N \to \infty$. The decreasing behavior of $C(k)$, when k is large, is due to the fact that a high degree vertex connected to vertices of much smaller degrees -precisely the situation of hubs in scale-free networks- inevitably exhibits small values of clustering coefficient regardless of it forms a part of many or a few triangles [Soffer and Vázquez, 2005]. The fact that hubs connect to vertices of moderate degrees implies that the formation of perfect cliques composed of vertices of moderate degree is usually not possible, which probably influences the location of the peak. Thus, the appearance of a peak in assortative random scale-free networks is natural; the precise form and position of the peak depends, however, on the particular form of the degree sequence of the network.

Finally, we examine the TS model. Figure 3.18 shows how the degree-dependent coefficient $C(k)$ behaves for different triangulated TS networks, all of them based on an original Barabási-Albert construction with $N = 3 \cdot 10^4$ and $m = 2$. Each curve of the picture results again from averaging over ten independent realizations of the algorithm. A notable characteristic of the curves is the "discontinuity" they present at $k = 24$. This feature of our TS networks, as we already saw, is a finite-size effect which would not appear in infinite versions ($N \to \infty$). Thus, let us focus on the part of the curves corresponding to $k < 24$. The figure shows that $C(k)$ approximately follows power laws like $C(k) \sim k^{\gamma}$ with $\gamma > 0$. This behavior proves to be exactly the opposite to that found in real dissortative networks, where $C(k)$ falls as power laws with negative exponents γ. The disagreement is evidently due to the fact that dissortative networks and TS constructions present very different topological properties.

The disagreement between the clustering coefficients of the models discussed and real networks is evidently due to the the fact that they present different topological properties. AS and DS models exhibit mean clustering coefficients many orders of magnitude smaller than those observed in real networks[9], and TS models do not reproduce the behavior $C(k) \sim k^{-\gamma}$ found in many dissortative networks. With respect to this, we want to emphasize that the goal of this

[9]In relation to this, we suggest a modification of the AS algorithm which perhaps could offer better values of C in the assortative case. Since vertices of small degree do not contribute appreciably to the mean clustering coefficient, one could apply the algorithm only when at least one of the four selected vertices at the corresponding step has a degree larger than any given k. Provided all values have a smaller degree, only the random step would be carry out. This procedure could lead to a larger value for C, as well as smaller average path lengths. In addition, it would be in better agreement with the real assortative networks, in which assortativity is only present between moderate-to-large-degree vertices.

chapter is to investigate the effect of correlations on the topology of networks, and not to construct models capable of reproducing the properties found in real networks. The algorithms are obviously mathematical constructions with the sole purpose of emulating certain desired properties (such as assortativity, dissortativity, and large transitivities), and not "realistic" algorithms designed to model networked systems. The topological properties of the models, however, confirm without a doubt the importance of vertex-pair correlations, and thus, they validate our approach.

Tomography. Tomography is a useful tool to investigate network topology. It deals with the study of the structure of layers which surround a given vertex in a network. The principal motivation for examining the tomography of a network results from its importance for understanding the spreading phenomena taking place in networks. We can ask for instance how a computer virus spreads in the Internet, or how information in a social network spreads from the original vertex to others. Spreading phenomena depend on how vertices and edges are distributed around the vertex from which the spreading starts. That is, spreading phenomena depend on the structure of layers around the initial node. Cohen et al. [2002b] examined the shells around the maximum-degree vertex in uncorrelated networks. Here, we analyze the tomography of our vertex-pair correlated networks. However, in this study we apply the procedure not only to the vertex with the maximum degree but to all vertices of the network. The procedure is similar to that which we use to measure the average path length of a network. We start from an arbitrary vertex (the root) and assign to it the shell number 0. Then, all edges starting at this vertex are followed and all vertices reached are assigned to shell number 1. Then, all edges leaving a vertex in shell 1 are followed and all vertices reached that do not belong to previous shells are labeled as vertices of shell 2. The same is carried out for shell 2, etc, until the whole network is exhausted. We then get $N_{\mathcal{L},r}(k)$ as the number of vertices of degree k in layer $\mathcal{L}$ for root r. The repetition of the whole procedure starting at all N vertices of the network gives us $P_{\mathcal{L}}(k)$, the degree distribution in shell $\mathcal{L}$. $P_{\mathcal{L}}(k)$ is then defined as

$$P_{\mathcal{L}}(k) = \frac{\sum_r N_{\mathcal{L},r}(k)}{\sum_{k,r} N_{\mathcal{L},r}(k)}. \tag{3.53}$$

We are actually most interested in the average degree $\langle k\rangle_{\mathcal{L}} = \sum_k kP_{\mathcal{L}}(k)$ of vertices of the layer $\mathcal{L}$. In the epidemiological context, this quantity can be interpreted as a disease multiplication factor after $\mathcal{L}$ steps of propagation. It describes how many neighbors a vertex can infect on average. Note that such a definition of $P_{\mathcal{L}}(k)$ gives us the following degree distribution in the first shell:

$$P_1(k) = \frac{\sum_r N_{1,r}(k)}{\sum_{k,r} N_{1,r}(k)} = \frac{kN_k}{\sum_k kN_k} = \frac{kP(k)}{\langle k\rangle}, \tag{3.54}$$

where, as always, $P(k)$ and N_k are, respectively, the degree distribution and degree sequence of the network. We bear in mind that every edge in the network is followed exactly once in each direction. Hence, we find that every vertex of degree k is counted exactly k times. From Eq. 3.54 it follows that $\langle k\rangle_1 = \langle k^2\rangle/\langle k\rangle$. This quantity plays an important role in the current percolation theory of networks [Cohen et al., 2000a] and depends only on the first and second moment of the degree distribution, but not on the correlations. Of course, $P_0(k) = P(k)$.

A similar study for more distant shells gets complicated because of the existence of correlations and cycles. Let us discuss for example the degree distribution in the second shell. In this case, we find that every edge leaving a vertex of degree n in shell 1 is counted $n-1$ times. Taking this into account, and neglecting the possibility of short cycles (which is appropriate in

the thermodynamic limit $N \to \infty$), we have:

$$P_2(k) = \frac{\sum_n nP(n)(n-1)P(k|n)}{\sum_n nP(n)(n-1)} \quad , \tag{3.55}$$

where $P(k|n)$ is the conditional probability that an edge leaving a vertex of degree n enters a vertex with degree k. Eq. 3.55 depends explicitly on the degree-degree correlations of the network through the term $P(k|n)$. In highly assortative networks, the conditional probability $P(k|n)$ is closely distributed around $n = k$, in such a way that $P^a(k|n) \sim \delta_{nk}$. Therefore, we can assume: $P_2^a(k) = \text{const} \cdot \sum_n nP(n)(n-1)P^a(k|n) \sim \text{const} \cdot kP(k)(k-1)$. Now, consider the corresponding uncorrelated network. The probability that a vertex of low degree l be found in shell 2 (that is, $P_2^r(l)$, with l small) is given by $\text{const} \cdot \sum_n nP(n)(n-1)P^r(l|n)$. Comparing this result with the result for the assortative network, $P_2^r(l) = \text{const} \cdot \sum_n nP(n)(n-1)P^r(l|n) > \text{const} \cdot \sum_n lP(l)(l-1)P^r(k|n) = \text{const} \cdot lP(l)(l-1) \sim P_2^a(l)$. The probability that a vertex of high degree h be found in shell 2 (that is, $P_2^r(h)$, with h large) is $P_2^r(h) = \text{const} \cdot \sum_n nP(n)(n-1)P^r(l|n) < \text{const} \cdot \sum_n hP(h)(h-1)P^r(k|n) = \text{const} \cdot hP(h)(h-1) \sim P_2^a(h)$. Thus, we expect that the degree distribution in shell 2 changes under assortative mixing in such a way that $P_2^a(k) > P_2^r(k)$ for k large, and $P_2^a(k) < P_2^r(k)$ for k small. Hence, we have: $\langle k \rangle_2^a > \langle k \rangle_2^r$. On the other hand, in strongly dissortative networks hubs tend to connect to vertices of minimum degree. Thus, when the degree considered h is large, $P_2^d(h) = \text{const} \cdot \sum_n nP(n)(n-1)P^d(h|n) \sim (\min)P(\min)(\min - 1)$, where min is the minimum degree of the network. Consider the corresponding uncorrelated network. The probability that a vertex of high degree h be found in shell 2 (that is, $P_2^r(h)$, with h small) is given by: $P_2^r(h) = \text{const} \cdot \sum_n nP(n)(n-1)P^r(h|n) > \text{const} \cdot \sum_n (\min)P(\min)(\min - 1) \sim P_2^d(h)$. We conclude that $P_2^d(k) < P_2^r(k)$ for k large. If the fraction of large degree vertices found in shell 2 is smaller than in uncorrelated networks, then the fraction of low degree vertices in shell 2 must be larger, so that the total number of vertices to be found in shell 2 is constant[10]: $k(k-1)N_k$. Consequently, $P_2^d(k) > P_2^r(k)$ for k small. From the preceding discussion results evident $\langle k \rangle_2^d < \langle k \rangle_2^r$.

Let us analyze the tomography of the AS networks. Fig. 3.19 corresponds to diverse AS-Barabási-Albert networks (all based on the same original Barabási-Albert construction of $N = 3 \cdot 10^4$ and $m = 2$), each one with a different degree of assortativity (see cmponents which lump together a non negligible fraction of vertices. Consequently, the giant component of an extaption of the figure). Panel (a) compares the shell structure of these networks, for which the assortativity ranges from fully uncorrelated ($\mathcal{A} = 0$) to strongly assortative[11] ($\mathcal{A} = 0.777$). The figure suggests that (regardless of the degree of the initial root) any spreading phenomenon on weakly assortative networks (a realistic case) will rapidly reach highly connected vertices, and then propagate slowly to vertices of dwindling degree. When the assortativity increases, the propagating agent does not reach the high-degree vertices so fast, and the spreading on distant shells, where the less connected vertices are found ($k = 2$), is still slower. Thus, the spreading agent infects the whole network more rapidly if the network is uncorrelated. Figure 3.19 (b) shows the cumulative distribution of the average number of vertices per shell

$$\bar{\mathcal{N}}(\mathcal{L}) = \frac{\sum_{i=0}^{\mathcal{L}} \sum_{r,k} N_{i,r}(k)}{\sum_{r,k,i} N_{i,r}(k)} \tag{3.56}$$

[10] In absence of short cycles.

[11] Note that tomography can only be investigated on fully connected networks, or on connected components. When the degree of assortativity is very large, initially connected AS networks break into separated coremely assortative network exhibits a different degree distribution $P(k)$ than the original network. For this reason, we give up comparing networks with degree of assortativity larger than $\mathcal{A} = 0.777$.

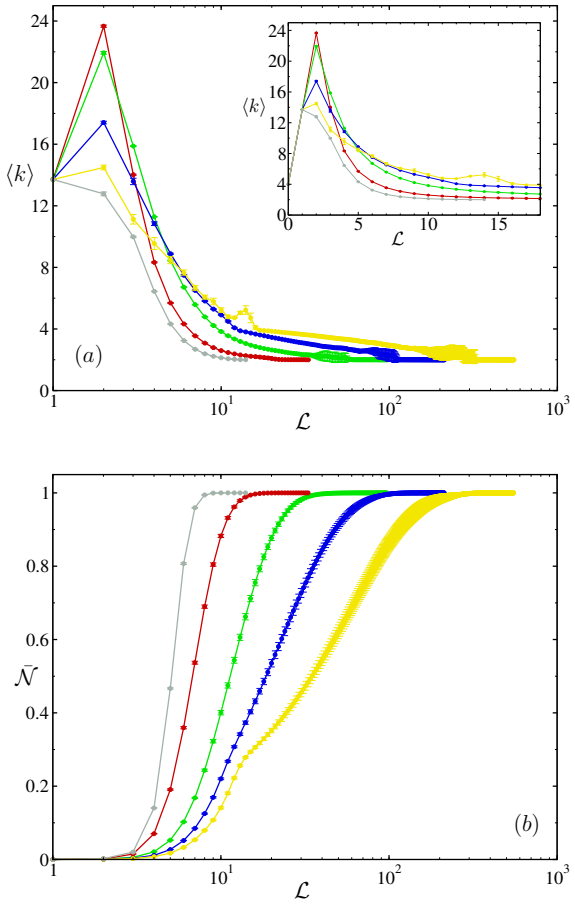

Figure 3.19: AS model. (a): Average degree, $\langle k \rangle_l$, as function of shell number $\mathcal{L}$. The curves correspond to the following values of assortativity: $\mathcal{A} = 0$ (grey), $\mathcal{A} = 0.221$ (rot), $\mathcal{A} = 0.443$ (green), $\mathcal{A} = 0.634$ (blue), and $\mathcal{A} = 0.777$ (yellow). Notice the logarithmic scale on the x-axis. The inset shows the same curves in linear scale, which allow us to observe the tomographical behavior on the shell $l = 0$. We can see from both pictures that $\langle k \rangle_1$ does not depend on correlations, as well as that the value of $\langle k \rangle_2$ decreases when assortativity grows (see text). (b): $\bar{\mathcal{N}}(\mathcal{L}) = (\sum_{i=0}^{\mathcal{L}} \sum_{r,k} N_{i,r}(k))/(\sum_{r,k,i} N_{i,r}(k))$ as function of shell number $\mathcal{L}$. The picture shows clearly that any spreading phenomenon will propagate more slowly the more assortative the network is.

as function of the shell number $\mathcal{L}$. It indicates that the diameter of an assortative network increases as the assortativity grows. The picture also shows another interesting feature of the assortative mixing: $\bar{\mathcal{N}}$ reaches the value $\bar{\mathcal{N}} \simeq 1$ in a number of shells $\mathcal{L}$ much smaller than the diameter of the network (note the logarithmic scale on the x-axis of the picture). Thus, in assortative networks most vertices concentrate on a relative small "area" (to the effect of few shells), where the maximum distance between two vertices is much smaller than the diameter of the network. Finally, we notice a quite abrupt change in the tomographical behavior when the assortativity increases, starting from uncorrelated networks ($\mathcal{A} = 0$) to slightly assortative ones (see again panel a). Numerical simulations detect a jump in the value of $\langle k \rangle_2$ with respect to uncorrelated ones (where $\langle k \rangle_2 \simeq \langle k \rangle_1 = \langle k^2 \rangle / \langle k \rangle$ in the thermodynamic limit); the value of $\langle k \rangle_2$ then decreases slowly as the assortativity grows.

Real networks present assortativity only among highly connected vertices, and probably, their global topology is also affected by metric effects related to the fact that they are located in a physical space. Thus, the tomographical properties of real networks could somewhat differ from the results for AS networks presented above. Note, however, that the general tendency of forming highly clustered communities of same-degree vertices, from which the previously described properties result, is a feature both of AS and assortative real networks. Hence, we are convinced that the properties of our assortative networks recreate in a global way the tendencies one would observe in real assortative networks. Taken together, we think that a disease will definitely reach highly connected vertices more rapidly in a real assortative mixed network than in an uncorrelated version of it, and that poorly connected vertices "are better hidden" in real

assortative networks than in uncorrelated ones. This agrees with various studies carried out on the topic of immunization on social networks, in which it is advised to vaccinate the highly connected vertices in order to prevent the propagation of diseases [Albert et al., 2000].

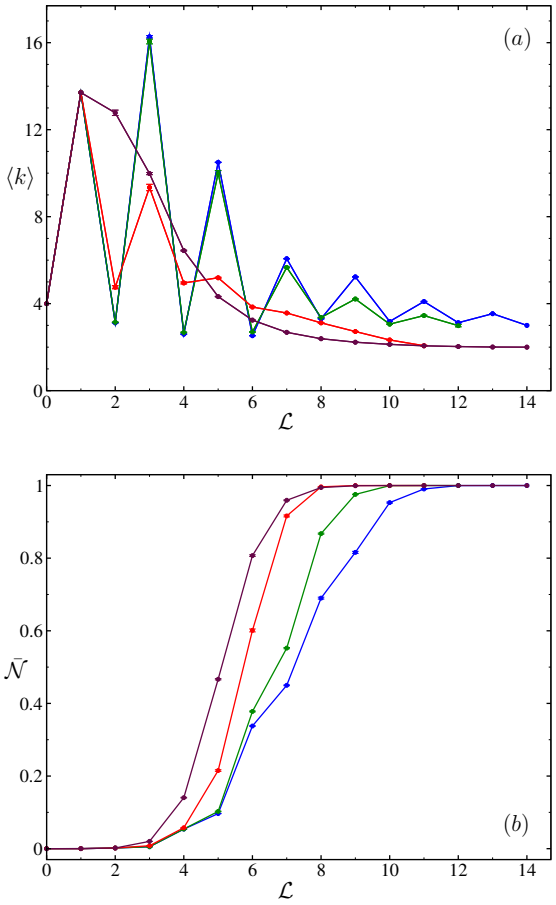

Figure 3.20: DS model. (a): Average degree, $\langle k\rangle_l$, as function of shell number $\mathcal{L}$. Here, we use again the parameter p to estimate the degree of dissortativity of the networks considered. From top to bottom (or from peaked to smooth): $p = 0$ (maroon), $p = 0.9$ (rot), $p = 0.999$ (dark green), and $p = 1$ (blue). The oscillating behavior is typical of dissortative networks (see text for details). Note also that the diameter is approximately the same as the diameter corresponding to the uncorrelated network (for some values of p, it is even a little smaller, which could indicate a major degree of compactness). (b): $\bar{\mathcal{N}}(\mathcal{L}) = (\sum_{i=0}^{\mathcal{L}} \sum_{r,k} N_{i,r}(k))/(\sum_{r,k,i} N_{i,r}(k))$ as function of shell number $\mathcal{L}$.

The study of tomography on DS-scale-free networks also yields interesting outcomes. As before, we use a Barabási-Albert network of $3 \cdot 10^4$ vertices and $6 \cdot 10^4$ edges ($m = 2$) as base-network to which we apply the DS algorithm. In the figure 3.20 (a) is plotted $\langle k\rangle_{\mathcal{L}}$ as function of the layer number $\mathcal{L}$. In uncorrelated (finite) networks the curve $\langle k\rangle_{\mathcal{L}}$ reaches a maximum value of $\langle k^2\rangle/\langle k\rangle$ for shell 1, and then it slowly decreases to $\langle k\rangle = 2$, the value which corresponds to the minimum vertex degree $k = 2$ -thus, the vertices lying nearly on the borders of the network are usually the poorly connected vertices-. In a dissortative network the behavior changes. Now, the value of $\langle k\rangle_{\mathcal{L}}$ oscillates as a function of $\mathcal{L}$; the second shell exhibits a local minimum of $\langle k\rangle_2$, followed by a maximum for $\mathcal{L} = 3$, and decreases again for shell $\mathcal{L} = 4$, etc. These jumps of the tomographical curve are typical for dissortative networks. The explanation is not complicated. As we have already seen, from Eq. 3.55 $\langle k\rangle_2 < \langle k\rangle_2^r$ must be satisfied. Now, if the second shell possesses a large number of low-degree vertices, the third shell must then be full of high degree vertices, because of the dissortative tendency of vertices to connect. For the same reason the shell $\mathcal{L} = 4$ must contain mostly vertices of small degree, etc. Thus, dissortative correlations produce networks where the propagating agent more readily reaches vertices of small degree than in uncorrelated ones. In these networks the lowly connected vertices do not form the "periphery" of the network, but rather play a more important role in the spreading phenomena, since they represent bridges between the highly connected ones. High-degree vertices are rapidly

reached too (see, for instance, the peaks at $\mathcal{L} = 3$ and $\mathcal{L} = 5$ of the figure), although alternating with low-degree vertices. The periphery of the network must then consist of a large fraction of medium-degree vertices, which are the last ones to be affected by the spreading agent.

Figure 3.20 (b) shows the cumulative distribution of the average number of vertices per shell $\bar{n}(\mathcal{L})$ as a function of shell number $\mathcal{L}$. Let us discuss it. We begin with the fact that the four plotted curves differ starting from the shell $\mathcal{L} = 3$, with $\bar{\mathcal{N}}(\mathcal{L} = 3)$ decreasing as the dissortativity increases. With respect to this, one must bear in mind that $\bar{\mathcal{N}}(\mathcal{L} = 2)$ is necessarily equal for all correlated networks (if short cycles may be neglected). To explain the different number of vertices in shell 3, we start from the already discussed fact that the fraction of low-degree vertices in shell 2 is larger under dissortative mixing than under uncorrelated mixing. However, vertices of small degree located at the second shell can only connect to a very small number of high-degree vertices in the third shell, because the number of edges which leave from a low-degree vertex is small. Starting from a smaller number of vertices in the third shell (in comparison to uncorrelated networks) hinders reaching a large number of further vertices and leads to a weaker population of the fourth shell, etc. The typical oscillating behavior of dissortative networks also affects the behavior of $\bar{\mathcal{N}}(\mathcal{L})$. A careful examination of the curves (particularly those that correspond to the more dissortative network) shows that the slope of straight lines connecting the points is smaller when passing from an even shell to an odd shell than when changing from an odd to an even shell. Note that this supports the statement that when changing from a shell with few hubs the next shell receives a smaller number of vertices than when passing from a shell with a large fraction of hubs. The simulation results unveil two other important properties of the dissortative DS networks: Irrespective of their degree of dissortativity, they always remain fully connected (our simulations show that even extreme dissortative networks ($p = 1$) are composed of only one component), and, in spite of the fact that the average path lengths are definitely larger than in uncorrelated networks, the diameter is always approximately the same.

To conclude, we discuss the tomography of the TS model. The study of shells of the TS networks gives in general similar results to those of assortative networks. This is not surprising, since triangulated networks are also somehow assortative networks. The behavior, however, exhibits some differences. Figure 3.21 (a) shows the average degree $\langle k \rangle_{\mathcal{L}}$ as a function of the shell number $\mathcal{L}$. As before, the results are averaged over ten realizations of the algorithm when applying it to a Barabási-Albert network of $3 \cdot 10^4$ vertices. Like in assortative AS networks, when the transitivity begins to increase, the mean degree of the second layer $\langle k \rangle_2$ undergoes an abrupt increment (it reaches the value $\langle k \rangle_2 = 20.155$ for $C = 0.0026$), which indicates a large number of high degree vertices in this shell; the average degree $\langle k \rangle_{\mathcal{L}}$ rapidly decreases for more distant layers. As the transitivity grows, the number of high degree vertices in the third shell increases (in comparison to the AS model). This is the natural consequence of the tendency of the algorithm to form cliques: When the network reaches their maximal transitivity (C_{max}), the network is composed exclusively of a certain quantity of cliques of diameter 1 (neglecting finite size effects). Networks which are strongly but not completely triangulated can be visualized as a set of highly triangulated clusters which are connected between themselves by a small number of edges; these highly triangulated clusters (which are actually quasi cliques) have diameter 2. (Here, the diameter of a cluster is the maximum distance between all vertex pairs of the cluster). On the other hand, strongly assortative AS networks are composed of highly connected clusters formed by all vertices having the same degree in the network; therefore, these clusters exhibit larger diameters than the quasi-cliques of the strongly triangulated TS networks. The point is that, in AS networks, vertices to be found in the third shell have usually the same degree as the root selected to start the tomographical procedure or a smaller one, if the root is a hub; since low-degree vertices dominate in scale-free networks, the average degree $\langle k \rangle_{\mathcal{L}=3}$ is not too large. In TS networks, however, the vertices having the same degree as the root are located

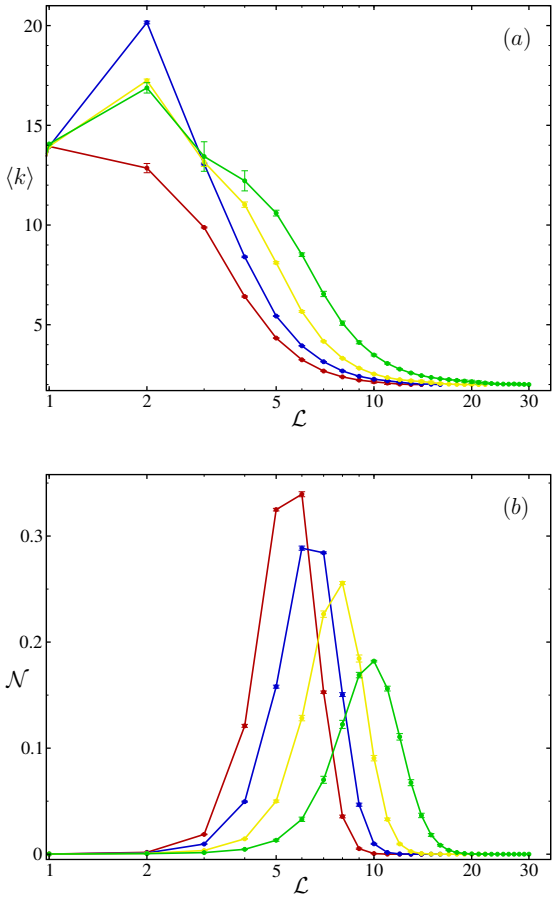

Figure 3.21: TS model (a): Average degree, $\langle k\rangle_l$, *versus* shell number $\mathcal{L}$. The curves correspond to different triangulated networks: $C = 0.0013$ (red), corresponding to the uncorrelated ones, $C = 0.0026$ (blue), $C = 0.0185$ (yellow), and $C = 0.1202$ (green). All networks were totally connected. For larger values of C the networks already break up into isolated components. Low triangulated networks show a behavior similar to that which we found for assortative networks. In strongly triangulated networks one can observe that even shells exhibit quite large values of $\langle k\rangle_{\mathcal{L}}$, clearly demonstrating the transitivization of the network (see text for details). (b): $\mathcal{N}(\mathcal{L}) = (\sum_{r,k} N_{\mathcal{L},r}(k))/N$ as function of shell number $\mathcal{L}$.

in shell 1 or shell 2 (because they all belong to the same quasi-clique), but vertices of shell 3 belong to different quasi-cliques than the one where the root is located. Since connections between quasi-cliques are not degree-degree dependent, vertices to be found in shell 3 can have any degree, which increases the value of $\langle k\rangle_{\mathcal{L}=3}$ in comparison to assortative networks.

Note that we restrict our analysis of shells to networks with $C \leq 0.12$, so that the networks considered are fully connected. Figure 3.21 (b) shows the mean number of vertices found in each layer $\mathcal{N}(\mathcal{L})$. (We do not use here the cumulative distribution, as above. Since the information provided by both distributions -the cumulative distribution and the one represented in figure 3.21- is the same, both can be interchangeably used. In this case, the representation chosen is more explicit). The behavior of $\mathcal{N}(\mathcal{L})$ shown in the picture is quite similar to that which we find for the assortative networks, and confirms the important increment of the average path length of networks exhibiting community structure.

Chapter 4

Percolation

Percolation theory describes the simplest possible phase transition with nontrivial critical behavior. It has found a huge variety of applications in many fields, from polymer modeling and particle diffusion in porous media to the estimation of oil reservoirs in rocky areas. In the past few years, robustness studies on the Internet and other networked systems, as well as new approaches in epidemiology, have also produced a burst of interest in percolation processes on networks. These studies, however, have hitherto been done on uncorrelated systems, and therefore, the predictions of such models might be inadequate for reproducing the percolation properties of real world networks. In this chapter we use the algorithms introduced in the preceding chapter to examine the influence of vertex-pair correlations on the percolation problem. We show that the concentration dependences of the giant component's order, as well as the existence and properties of the percolation threshold, are strongly different for uncorrelated and correlated networks.

4.1 The percolation problem

Percolation is certainly one of the best studied problems in statistical physics. It deals with the interconnection of a given set of points randomly positioned in a space. The classical percolation space is a regular periodic graph embedded in $\Re^n$, for $n < \infty$ (see figure 4.1). These types of graphs are called lattices in physics, and they are typically used to model the structure of crystals (materials that have an orderly and periodic arrangement of atoms in three-dimensional space). The vertices of a lattice -the intersections of lines in the figure- are called sites, and the edges -the lines linking two intersections- are called bonds. Let us now assume that all sites are assigned to one of two classes: to the class of *occupied* sites (with probability p), and to the class of *vacant* or *non occupied* sites (with probability $1-p$). Of course, one may replace "occupied" and "vacant" by another pair of words denoting mutually exclusive states of the site. The crucial requirement is the independence of the assignment, that is, all sites must be assigned regardless of the state of the other sites. In this space, where each site is either occupied or not, neighboring sites belonging to the same class form components, so that components composed of sites of one of the classes are separated from each other by sites of the another class. The goal of percolation theory is to study the number and properties of these components or clusters, as they are usually called in the percolation theory.

There are many types of lattices, which differ by the dimension of the space in which they are embedded, and the structure of interconnections. Thus, examples of two-dimensional lattices are: the *square* lattice, the *triangular* lattice, the *honeycomb* lattice, etc. In physics we can find many well-known lattices of three dimensions: the *simple cubic* lattice, the *body-centered* lattice,

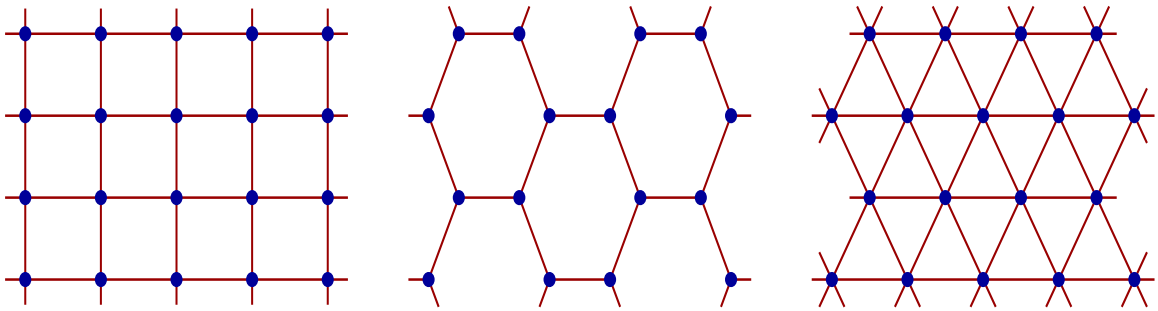

Figure 4.1: Definition of square, honeycomb, and triangular lattices, respectively. The smallest volume in a lattice that represents one full repetition of its periodic vertex arrangement is called a unit cell. The manner in which the vertices are arranged in a unit cell is known, in physical terms, as its crystal structure. The lengths of the edges of a unit cell along its major axes are known as lattice constants, which are usually denoted by three numbers, a, b, and c. In some crystal structures, however, the edge lengths along all axes are equal ($a = b = c$), so only one lattice constant is used for its dimensional description, a.

the *face-centered* lattice, the *diamond* lattice, etc. The percolation properties of a lattice depend on the particular topology of the lattice, given that the size and shapes of components depend not only on the probability p with which the classes are assigned, but also on the dimension and wiring structure of the lattice. Nevertheless, the same qualitative behavior can always be observed independent of the type of lattice: when the concentration of occupied (vacant) sites is small, the system consists of many infinitesimal components of occupied (vacant) sites, however, when the concentration is large enough, a giant component exists (typically one only) which extends from one side to the opposite side of the lattice. In this last case, it is said that the component percolates through the system.

Everything we have defined till now is related to the so-called "site percolation". We can also consider processes of bond percolation, in which bonds (instead of sites) are assigned to one of two mutually exclusive classes: the class of occupied bonds (with probability p), and the class of non occupied bonds (with probability $1 - p$). The two kinds of bonds are also called *passable* or *open* bonds and *blocked* or *closed* bonds, respectively, in related studies. Thus, in "bond percolation" a lattice contains occupied and non-occupied bonds randomly distributed in the lattice. The study of the emerging components, however, is done under two different perspectives: i) where components are groups of neighboring occupied (non occupied) bonds [Hsu and Huang, 1999], whereby two neighboring bonds are bonds which are incident to the same site, and ii) where components were groups of neighboring sites, whereby two sites are considered neighbors when they are connected by occupied (non-occupied) bonds. In both cases, the system shows a behavior qualitatively similar to (though different in some details from) site percolation. As p grows the order of the numerous small components increases, so that at a percolation threshold p_c a spanning component appears. However, the point p_c at which the percolating component appears for the first time differs for bond and site percolation. In fact, Hammersley [1961] proved that the percolation threshold for site percolation is always larger or equal to that for bond percolation.

Site and bond percolation have been found useful to characterize many disordered systems, such as porous media [de Gennes and Guyon, 1978, Sahimi, 1994, Machta, 1991, Moon, 1995], granular materials [Hufnagel et al., 1999, Odagaki and Toyofuku, 1998], fragmentation and frac-

tures [Campi et al., 2000], polymers [Bunde et al., 1995], gelation [Lairez et al., 1991], random-resistor-insulator systems [de Arcangelis et al., 1985], dispersed ionic conductors [Roman, 1990], forest fires [Henley, 1993], biological evolution [Ray and Jan, 1994], thermal critical phenomena [Stanley, 1971], epidemics [Cohen et al., 2000a, 2002a] and robustness of networks [Moore and Newman, 2000, Newman and Watts, 1999b, Callaway et al., 2000]. The pure geometrical nature of this transition and its compelling application to diverse physical problems have drawn the attention of many researchers and turned percolation theory into a classical field [Stauffer and Aharony, 1994, Shante and Kirkpatrick, 1971, Essam, 1972, Isichenko, 1992, Kirkpatrick, 1973, Kesten, 1982, Zallen, 1983, Sokolov, 1986, Stauffer, 1979]. It is one of the simplest and best understood examples of a phase transition, but there are however many things about it that are still not known. For example, despite decades of considerable effort [Stauffer, 1979, Essam, 1980, Hammersley and Welsh, 1980, Wierman, 1982], no exact solution of the site percolation problem yet exists on the simplest two-dimensional lattice, the square lattice, and no exact results are known for any lattice in three or higher dimensions. (For a review of the few mathematical results which have been rigorously proven see [Kesten, 1982]). Because of these and many other gaps in the current understanding of percolation, numerical simulations have found wide use in the field.

The threshold of percolation (both for site and for bond percolation) depends not only on the type of lattice but also on its size. In general, one does not have a sharply defined threshold in finite lattices. The concentration p_c at which the percolating cluster for the first time appears depends on the size of the lattice. However, one can observe that p_c tends to a concrete value as the size of the lattice increases, and that the appearance of the spanning cluster somehow becomes more and more sudden. This does not sound totally unfamiliar to physicists who are already acquainted with *phase transitions* and *critical phenomena*. The paradigm of a continuous phase transition in physics is the conversion at the *Curie temperature* $T_c = 1043K$ of iron from paramagnetic to ferromagnetic form. Iron, like copper or zinc, is paramagnetic at $T > T_c$ (not magnetized in absence of an applied magnetic field), while for $T < T_c$ the material is ferromagnetic (magnetized even when no magnetic field is applied). The iron magnetization decreases steadily as T_c is approached, vanishing entirely at T_c and all higher temperatures. Additionally, the rate of change of magnetization changes discontinuously at T_c. Other examples of phase transitions in physical systems are critical opalescence, liquid-gas transition, helium I/helium II transition, conductor/superconductor transitions, etc [Stanley, 1971, Binney et al., 1992]. The percolation threshold is also a phase transition. It is defined as the probability p_c at which a percolating component appears for the first time in an infinite lattice. For $p > p_c$, a spanning component extends from one side of the system to the other, whereas for all $p < p_c$ no such cluster exists. Any "effective" threshold value obtained numerically or experimentally on finite lattices needs thus to be extrapolated carefully to infinite system size (the thermodynamic limit). The theory attempting to deal with all aspects of critical phenomena, and, in particular, to extrapolate the point at which phase transitions occur, is the so-called scaling theory, which uses techniques of the renormalization group theory [Wilson, 1982, Binney et al., 1992].

Historically, percolation theory goes back to Flory and to Stockmayer, who, during the second World War, used it to describe how small branching molecules form larger and larger macromolecules if more and more chemical bonds are formed between the original molecules. This polymerization process may lead to gelation, that is, to the formation of a network of chemical bonds spanning the whole system. Thus, the original small molecules correspond to our sites, the macromolecules to our clusters, and the network to our percolating cluster. Flory and Stockmayer used a Cayley tree (a Bethe lattice in physical terms) on which they developed the theory that is today called the percolation theory. The beginning of the percolation theory is however usually associated with a publication from Broadbent and Hammersley from 1957, which

introduced its name and dealt with it more mathematically, using geometrical and probabilistic concepts [Broadbent, 1954, Broadbent and Hammersley, 1957, Hammersley, 1957]. The example that they discussed was a model for the spread of fluid or gas through a random medium. The fluid, for example, spreads through channels, moving through a channel if and only if the channel is wide enough. There is therefore no randomness in the motion of the fluid itself -such as in a diffusion process- but only in the medium, i. e., in the system of channels. They modeled this system of channels and intersections of channels as a lattice. The channels were the edges or bonds between adjacent sites, i. e., the intersections. Each bond was passable (blocked) with probability p $(1-p)$, while the state of a bond was independent of the state of the others. The question was whether the fluid from outside a large region could reach the origin.

Percolation can also be used as an idealized simple model for the distribution of oil or gas inside porous rocks in oil reservoirs. Imagine that unoccupied positions represent regions filled with hard rock, while the occupied positions represent pores that are filled with oil or gas. In real reservoirs, the physical mechanisms that created oil deposits imply some correlations between pores due to the way the rock was cracked originally or the placing of the different deposits. The simple percolation model ignores these correlations and assumes that each basic point is solid rock or pore independent of its neighbors. However, the qualitative features hold in the more realistic models. When the concentration of pores is small, the oil is only found in small connected clusters (components). Therefore, if we place a well at a random position, it will most likely hit a small cluster, producing a small amount of oil, and hence be a bad investment. To produce a large amount of oil, a reservoir is needed where the concentration of pores -the porosity- is larger than a certain critical porosity, so that a giant cluster is more likely to exist.

Another example is the diffusion of hydrogen atoms through many solids, an effect which might be important for energy storage. If the solid is not a regular lattice, this diffusion takes place in a disordered medium. A simple disordered medium is a percolation lattice, where only a fraction p of all positions are occupied, and the rest is empty. Let us assume the hydrogen atom can move only from one occupied site of the lattice to a nearest neighbor which is also occupied. Then, the motion is respectively restricted to the cluster to which the atom initially belongs; it can never jump to another cluster. Moreover, suppose that the hydrogen atom moves randomly through this lattice selecting in each position one of the other possible neighboring positions. After a certain time t, one can calculate the average of squared distance R between the starting point and the end point. Diffusion on a regular lattice without disorder gives $R \propto t^k$, with $k = 1/2$. In a percolation lattice, it can be observed that for a concentration far beyond p_c, k is close to (but smaller than) $1/2$ for large times, whereas for concentrations far beneath p_c the distance R approaches a constant for large times, that is $k = 0$. The effect that for concentrations above p_c, the exponent k is neither this for normal diffusion ($k = 1/2$) nor for a constant distance ($k = 0$), is called anomalous diffusion.

Classical percolation makes use of lattices as spaces for percolation models. However, general graphs can also be used. Percolation processes have been recently used to study the robustness of networks against random breakdowns of (or attacks to) their elements. When networks are subject to random breakdowns, that is, when a fraction q of vertices -together with their connections- is removed randomly, their integrity might be compromised. Vertices of the network are classified again as non-occupied vertices if they are removed and occupied vertices if they remain in the system after the elimination process. When q exceeds a certain threshold, $q > q_c$, the network disintegrates into smaller, disconnected components. Below the critical threshold q_c, a giant component composed exclusively of occupied vertices still exists, whose order is proportional to that of the entire system. Note, however, that since one cannot define sides or other metrical properties in networks, the percolating component must be defined in a different way: A component is said to be a percolating or spanning component (in an infinite network)

if the fraction between the number of occupied vertices forming the largest component and the order of the whole network is larger than zero, which also generalizes a property that satisfies the percolating components on infinite lattices [Kesten, 1982]. The percolation threshold q_c is thus defined only for infinite networks, and corresponds to that point at which $\mathcal{M}$ vanishes for the first time. Of course, the existence and the characteristics of the percolation threshold depend strongly on the particular topology of the network, just like the threshold depends on the type and dimension of lattices in classical percolation.

Percolation network models have also been used successfully in the field of epidemiology. Since it has become clear that many effects can only be described taking into account that an infecting agent does not spread in a homogeneous population of susceptible individuals, but over a network of contacts [Sander et al., 2002, 2003, Hufnagel et al., 2004], networks have been of great importance in studies on models of infection spread. In many cases it is important to know exactly the network structure, in other situations one can rely on simple model assumptions intended to mirror the properties of the system. In such epidemiological models, vertices are divided in immunized and non-immunized vertices. If the concentration of immunized vertices q is small, then the subnetwork formed by the non-immunized vertices consists basically of one giant component; however, if the concentration q is large, the subnetwork is exclusively composed of numerous small components. Similar to a hydrogen atom in a disordered medium, an infection could spread through this subnetwork of non-immunized vertices and if q is small affect a large fraction of them; if the fraction q is large enough, the disease can only propagate inside the small component which the originally infected vertex belongs to, but not to other non-immunized vertices. Therefore, the disease can only affect a very small fraction of the population if q is large. In this scenario, the question is which minimum fraction of vertices must be randomly immunized so that the infection does not affect more than a minority of non-immunized vertices. In terms of the percolation theory, the question is equivalent to that of finding the percolation threshold q_c for the networked system[1] [Eguíluz and Klemm, 2002, Pastor-Satorras and Vespignani, 2001, Vázquez and Moreno, 2003].

Phase transitions on random graphs were already studied in the late 1950's and early 60's by Erdös and Rényi. Their work is one of the few theoretical studies on phase transitions on not embedded graphs, whose results could be proven in depth [Erdös and Réyi, 1959]. Recently, percolation processes have also been studied on small-world [Moore and Newman, 2000, Newman, 2003a], and scale-free [Albert et al., 2000, Newman and Ziff, 2001] networks, although the results have not been demonstrated extensively from a mathematical viewpoint. Cohen and collaborators Cohen et al. [2000a, 2002a] present a heuristic argument indicating that uncorrelated scale-free networks with diverging $\langle k^2 \rangle$ -second moment of the degree distribution- undergo no transition under random elimination of sites or bonds, which means that a component spanning the system survives even for arbitrarily large fractions of crashed vertices ($q_c = 1$). On the other hand, Vázquez and Moreno [2003] proposed heuristic arguments suggesting that assortative networks may not exhibit a percolation transition even when the second moment of the degree distribution is finite. As far as we know, there are no further studies related to percolation

[1] This is not so simple in reality because social networks are dynamical systems. Assume that at time t there is a small fraction of non-immunized vertices in the network which are distributed into many small components. Given the dynamical character of the system, disconnected components at time t could form only one component thereafter, or this component could break into other smaller components for subsequent times. If the propagation and subsequent extinction of the disease occurs so rapidly that, in this period of time, the component containing the infected vertex does not come into contact with other components, then the infection cannot spread, and only vertices of the originally infected component can become infected. But if the infected component establishes contact with other components before the disease dies out, it could by all means propagate throughout the whole system. Thus, the dynamical aspect of the problem makes it more complex, and, therefore, it must be examined under the corresponding dynamical perspective.

processes on correlated networks.

The goal of this chapter is to study the percolation properties of vertex-pair correlated networks. We make use of the models we presented in the previous chapter in order to study the influence of degree-degree correlations and transitivity on the percolation features of networks. In particular, we focus specially on the behavior near q_c, in order to establish the existence (or not) of the percolation transition in correlated networks. Due to the great difficulties of a theoretical approach to the network percolation problem, we resign to investigate the problem numerically. Thus, extensive numerical simulations on scale-free networks demonstrate that correlations modify enormously the percolation behavior: Dissortative scale-free DS networks show a true percolation threshold irrespective of the finiteness or infiniteness of $\langle k^2 \rangle$. Assortative scale-free AS networks seem to have no percolation transitions, even when $\langle k^2 \rangle$ is finite, as Vázquez and Moreno [2003] suggest. Highly triangulated TS networks exhibit phase transitions even when $\langle k^2 \rangle$ is infinite.

4.2 Typical quantities and approaches

At the phase transition, a system changes its behavior qualitatively for one particular value of a continuously varying parameter. In the case of a percolation process on a lattice, if p increases smoothly from zero to unity, then we have no percolating cluster for $p < p_c$ and a unique [Harris, 1960] percolating cluster for $p > p_c$. When $p = p_c$, and only then, something peculiar happens: for the first time a path connects two opposite sides of the lattice. In the limit of an infinite lattice $L \to \infty$, the number of sites belonging to the percolating cluster is infinite (the reason being, if a path connecting two opposite sides of an infinite lattice exists, then this path must inevitably contain an infinite number of sites). On the other hand, below the critical p_c all components contain a finite number of sites, extending finitely in all directions of the $\Re^n$ space. Therefore, in order to establish the concentration p at which the phase transition takes place, an interesting quantity to be defined is the probability of which a given site is a member of the percolating cluster

$$P_\infty(p) = \frac{\text{number of sites in the percolating cluster}}{\text{total number of lattice sites}} \ . \tag{4.1}$$

This quantity obviously satisfy that $P_\infty(0) = 0$, $P_\infty(1) = 1$ and that P_∞ is non-decreasing. The critical probability p_c can be found as

$$p_c := \sup\{p : P_\infty = 0\} \ , \tag{4.2}$$

and is such that $P_\infty(p) = 0$ for $p < p_c$, and $P_\infty(p) > 0$ for $p > p_c$. This quantity, $P_\infty(p)$, is usually called the *strength* or *mass* of the infinite cluster. (Some authors define another similar quantity, namely, the probability that a given occupied site belongs to the percolating cluster).

Another important quantity is the number of s-clusters per unit site, n_s, where s-clusters are simply components formed by s occupied sites. (Following the classical nomenclature used in the percolation theory, we use the term *cluster* as synonymous to component.) It is given by

$$n_s(p) = \frac{\text{number of clusters of order } s}{\text{total number of lattice sites}} \ . \tag{4.3}$$

Thus, the probability that a site (occupied or not) belongs to a cluster of order s is sn_s. Spanning clusters are eliminated from $n_s(p)$, so that $P_\infty + \sum_s s n_s = p$, where the sum includes all finite components s and excludes the infinite cluster. This equation simply states that all occupied sites either belong to the infinite component, with probability P_∞, or to one of the finite clusters,

with probability $\sum_s sn_s$ (since single occupied sites surrounded by vacant neighbors are also considered clusters of order unity). We can now define $w_s = sn_s/(\sum_s sn_s)$, i. e., the probability that the finite cluster, to which an arbitrary occupied site belongs, contains exactly s sites. Hence, the so-called *mean cluster size* is given by:

$$S = \sum_s sw_s = \frac{\sum_s s^2 n_s}{\sum_s sn_s} \ . \tag{4.4}$$

As before, spanning clusters are eliminated from the sum[2].

P_∞ and S are two of the more important quantities used in the percolation theory to estimate the critical probability p_c at which the phase transition occurs. The fact that exact theoretical solutions are extremely hard to obtain, the percolation problem is typically approached by studying the asymptotic behavior of both quantities as p approaches the critical probability p_c. Analogous to results in statistical mechanics (in particular, to those related to critical phenomena), when fitting the asymptotic behavior, the emphasis in the physics literature has shifted to power laws or scaling laws. Extensive numerical evidence supports the assumption that the singular behavior equals powers of $|p - p_c|$ [Stauffer, 1979, Essam, 1980]. More specifically, one expects that

$$P_\infty \sim (p - p_c)^\beta \ \ , \ \ p \to p_c^+ \qquad \text{and} \qquad S \sim |p - p_c|^{-\gamma_\pm} \ \ , \ \ p \to p_c^\pm \tag{4.5}$$

for suitable constants $0 < \beta, \gamma_\pm < \infty$, where the plus (minus) refers to the approach of $p - p_c$ to zero from the positive (negative) side. It is also conjectured that the so-called critical exponents $\beta, \gamma_\pm$ are universal, that is, that their values do not depend on the detailed structure of the lattice, but on its dimension only. In other words, these exponents should be the same for all periodic graphs embedded in $\Re^d$ with one particular d. Numerical evidence presently available does not seem to rule this out. However, from a mathematical viewpoint, Kesten [1982] points out that power laws as in Eq. 4.5 have been established for very few models, and not at all for any of the extensively studied percolation models. (For a discussion of power law estimates see [Kesten, 1982].)

The important region for critical exponents is the region around p_c. Sadly, this is also the region where the effects of lattice finiteness are most visible. The question is thus how the quantities of interest behave near the percolation threshold in large but finite lattices. To deal with this question the *finite-size scaling* formalism Stauffer and Aharony [1994] makes use of another quantity called the correlation length ξ. One also assumes that ξ diverges with a power law as p approaches p_c:

$$\xi(p) \sim L \sim |p - p_c|^{-\nu} \qquad \text{, and inverting} \qquad |p - p_c| \sim L^{-1/\nu} \ . \tag{4.6}$$

The principal result of this scaling law is that the difference between the effective threshold p_{ef} (point at which a spanning cluster appears in the finite lattice for first time) and the percolation threshold p_c (for the infinite lattice) scales as $p_{ef} - p_c \sim L^{-1/\nu}$. Other finite-size scaling relations easy to obtain from Eq. 4.6 are $P_\infty(p_c, L) \sim L^{-\beta/\nu}$ and $S(p_c, L) \sim L^{\gamma/\nu}$. All this can be used to determine the critical exponents ν, β, and γ and the percolation threshold p_c of the system, while only measuring p_{ef} for different sizes L by means of computer simulations.

Another approach currently used to deal with the problem is to make use of *renormalization group* techniques [Kadanoff, 1966, Wilson and Kogut, 1974], which were first applied to

[2] Another possible definition could be $\sum_s sn_s / \sum n_s$. This is the average cluster size if every component -and not every site as in Eq. 4.4- is selected with equal probability.

thermal phase transitions and only afterwards to percolation. It attempts to justify the scaling assumptions commonly employed in the theory of critical phenomena, and to calculate the critical exponents entering these scaling assumptions. These techniques can be roughly divided into two classes: *field-theoretical* techniques, which were developed by pursuing the analogy between statistical mechanics and quantum field theory, and *real-space renormalization* techniques, applicable only to models based on lattices, involving quantities dependent on position coordinates in ordinary space. For percolation processes, the real-space renormalization is evidently the appropriate approach. However, this renormalization can only be applied to lattices which have a "discrete scaling symmetry", that is, which satisfy the following conditions: i) It must be possible to divide the sites into groups or *blocks*, and then to replace each block by just one single "super-site" ii) The new lattice must exactly resemble the one we started with, except for an increase in the lattice parameter (see caption of figure 4.1) by a factor b depending on the "blocking" process. A super-site is said to be occupied (vacant) if a cluster of occupied (non-occupied) elements percolates the block. If in the old lattice a site was occupied with probability p, then the probability p' that a super-site belonging to the renormalized lattice is occupied will in general be different and depend on the realized division in blocks. The basic idea of renormalization is *self-similarity* from the critical point p_c on. The word self-similar means that a phenomenon reproduces itself on different time and/or space scales. (More concretely, physical phenomena are called *similar* if they differ only in respect of the numerical values of the dimensional governing parameters [Barenblatt, 2003]). Thus, self-similarity assumes that the system will be similar for all space scales considered from the critical point $p' = p = p_c$ on. Taking this into account, one can obtain the percolation threshold from the equation $p'(p) = p$ where $p'(p)$ relates to both probabilities p' and p. Moreover, since the correlation length ξ limits the validity of similarity [Stauffer and Aharony, 1994], ξ must be the same in both the original lattice and the renormalized lattice of super-sites. If in the original lattice we have $\xi = \text{const} \cdot |p - p_c|^{-\nu}$ then in the renormalized lattice we have $\xi' = \text{const} \cdot |p' - p_c|^{-\nu}$, with the same proportionality constant and the same critical exponent ν, provided that both $|p - p_c|$ and $|p' - p_c|$ remain very small. However, the new lattice has a new lattice constant b, and ξ' must be measured in these units. Hence, $\xi' = \xi/b$, and

$$b|p' - p_c|^{-\nu} = |p - p_c|^{-\nu} \ , \tag{4.7}$$

which is the basic equation of real-space renormalization. In this way, it is also possible to obtain the value of the critical exponent ν. Taking the logarithm of both sides, one yields

$$\frac{1}{\nu} = \frac{\log[(p' - p_c)/(p - p_c)]}{\log b} = \frac{\log \lambda}{\log b} \tag{4.8}$$

where $\lambda = (p' - p_c)/(p - p_c) = dp'/dp$ at $p = p_c$. Simple examples of real-space renormalization on lattices, both in node and bond percolation, can be found in [Stauffer and Aharony, 1994, Binney et al., 1992].

Unfortunately, none of the preceding classical techniques can be applied when studying the percolation behavior of networks. There are two reasons for this: networks are usually not embedded in any space (bear in mind that lattices are regular periodic graphs embedded in $\Re^n$), and, moreover, they present complex degree distribution functions and vertex-pair correlations. The fact that networks are embedded in non-metrical space implies that they are topological constructions in which no metrical concepts are defined. In particular, vertices are not placed in a given spot of $\Re^n$ (as happens in lattices), and, consequently, "opposite sides of the network" is a notion which in this case has no meaning. We are thus compelled to redefine the notion of "percolating component" when dealing with network percolation processes, but it is sadly not an easy task when applied to non embedded finite networks. Therefore, applying the finite-size

scaling formalism -which requires to know the probabilities $p_e f$ at which the system percolates when it is finite- is impossible. On the other hand, the fact that networks are usually correlated and present no regular degree sequences makes it very difficult to implement the renormalization approach: designing an appropriate blocking process for the general case is an enormous task.

The network percolation therefore presents important methodological differences in comparison to the lattice percolation. The percolation process is basically the same: Each vertex (edge) is assigned to be occupied with probability p, or non occupied with probability $1-p$, regardless of the status of the other vertices (edges) in the network. In network percolation, however, it is common to describe the subnetwork of occupied (non-removed) vertices (edges) in terms of the fraction $q = 1 - p$ of non-occupied (removed) vertices (edges). However, the techniques and approaches which are usually employed for investigating lattice percolation processes cannot be used in general to study the percolation behavior of networks. The fact that networks are topological entities where no metrical concepts may be defined compel us to focus on the study of network percolation from a different perspective.

4.3 Phase transitions in lattices and random graphs

The percolation properties of random graphs [Erdös and Réyi, 1959], and some classical lattices [Stauffer, 1979], were already established a long time ago. In this section we reproduce numerically some of their principal results. For this purpose, we use new ways to approach the issue, which we will later apply in order to approach the general network percolation problem. Our aim is twofold: on the one hand, we aim to verify the validity of the new quantities we will introduce (since they must evidently be capable of reproducing the same results known for random graphs and lattices), and, on the other hand, we seek to check our simulation programs (since the study will be completely numerical).

Non embedded regular periodic graphs. Let us consider a 4-regular periodic graph with periodic boundary conditions (figure 4.2). This network evidently has the same local structure than the square lattice, and, in effect, both graphs are topologically identical in the thermodynamic limit. Hence, we expect that a percolation transition takes place at the same p_c as in the square lattice. Let us investigate its percolation behavior from just the topological viewpoint, without considering metrical aspects such as "borders", "frontiers", or "sides".

The first natural question is how the notion of percolating component can be defined for networks where the concept "sides of a network" makes no sense. In lattices, the spanning component connects not only two opposite sides of the lattice, but also contains an infinite number of vertices. Also, before the system percolates, all components of the lattice are finite (a spanning cluster does not exist). This feature of the lattice percolation is exploited by quantity P_∞ to characterize the phase transition. However, this is valid only for infinite networks. Another possible definition could be based on the quantitative measurement of the diameter of clusters. It could be defined that a component percolates through the system when its diameter is equal to the diameter of the network. This definition could be appropriate for regular graphs, but unfortunately it is unreasonable for more general networks. Bear in mind, for example, that the typical diameter of assortative networks is much larger than the "area" in which most of the vertices concentrate in these networks (remember the discussion related to figure 3.19 *b*). In the assortative case, a giant component containing a large fraction of vertices always appears long before it acquires the diameter of the network. This is obviously undesirable, because the practical uses of network percolation are focused precisely to obtain the critical probability p_c at which a non-negligible fraction of vertices is forming one single component. On

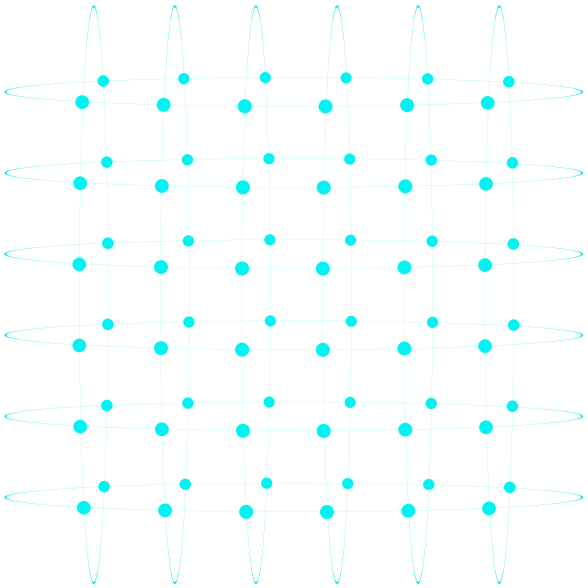

Figure 4.2: Example of a small 4-regular periodic graph, which is not embedded in any metrical space. The network contains 72 vertices and 144 edges. Note that this type of network has the same local structure than the square lattice, but that it lacks borders or any type of frontiers. In the thermodynamic limit the structure of this network is the same than that of the square lattice.

the other hand, this proposal suffers from the practical disadvantage that measuring diameters of large components -most notably when N is large- needs a lot of computational time. To recapitulate, the problem is that no appropriate definition of the percolation transition exists for finite networks.

The unfortunate situation is the following: On the one hand, it is possible to investigate percolation processes by means of simulations on finite networks only, because only systems having a finite number of elements can be numerically simulated. On the other hand, phase transitions in a finite network cannot be distinctly detected because the system undergoes no qualitative change as the probability p increases. (If a qualitative change happened, then it could be used to explicitly define when the phase transition occurs.) That does not mean, however, that the system does not undergo any abrupt change which can be related to phase transitions with varying p. In some small interval of p, the system experiences certainly an abrupt change in the percolation behavior (quantitatively speaking). But the exact point, within this interval, at which the behavior change occurs cannot usually be accurately established.

Consider again the quantities P_∞ and S. They cannot be used to deal with the problem since they are in need of a precise definition of the phase transition in order to determine when the percolating component appears for the first time in the system. Thus, we are somewhat obliged to generalize these quantities and introduce new indicators for detecting phase transitions in networks. For this reason, [Cohen et al., 2000b] propose the following quantity:

$$\mathcal{M}(q) = \frac{\text{number of non removed vertices in the largest component}}{\text{total number of vertices in the network}} \,, \tag{4.9}$$

which is probably the most simple generalization of P_∞. We will refer to it as the *vertex mass* of the largest component. Consider now the system when q, the fraction of removed vertices, is small or large. When q is small, we expect that the (initially fully connected) network still contains a giant component, in which almost all non-removed vertices are concentrated. This giant component will be the largest component in the system, and, in the thermodynamic limit, contain an infinite number of vertices. The vertex mass $\mathcal{M}$ will be strictly positive. When q is large enough, the giant component will break up into small finite clusters, even when $N \to \infty$. The fraction $\mathcal{M}$ will then vanish. Therefore, as q increases, the initially positive values of the

vertex mass, $\mathcal{M}$, will vanish at a certain critical concentration q_c, where, by definition, the phase transition takes place. Another similar quantity can be defined. Consider, for example, that we take into account the number of edges[3] in the largest component, instead of the number of vertices. (Note that this can be done when the percolation process corresponds to site percolation or bond percolation). The counting of edges, instead of vertices, will give the same percolation threshold q_c in the thermodynamic limit. The proof is evident: If the largest component contains infinite vertices, then it must also contain an infinite number of edges because the number of edges of a component is at least equal to its order minus one. On the other hand, if the size of the component is infinite, then its order must also be infinite, because, by reductio ad absurdum, if the component's order were finite, say N_c, then their maximum size (have in mind that our networks are simple graphs) would be $N_c(N_c-1)/2$, which is a finite quantity. For the same reasons, if the order of the largest component is finite, the size must inevitably be finite, and vice versa. Thus, another quantity we can introduce is the *edge mass* of the largest cluster

$$\bar{\mathcal{M}}(q) = \frac{\text{number of passable edges in the largest component}}{\text{total number of edges in the network}} \ . \tag{4.10}$$

Both quantities exhibit properties similar to ones of P_∞. It is obvious that in the limit of an infinite network $\mathcal{M}(1) = 0 = \bar{\mathcal{M}}(1)$; if the initial network is fully connected, $\mathcal{M}(0) = 1$ and $\bar{\mathcal{M}}(0) = \langle k \rangle$; and both $\mathcal{M}$ and $\bar{\mathcal{M}}$ are non-increasing. The critical probability q_c can then be obtained from the conditions

$$q_c := \inf\{q : \mathcal{M}(q) = 0\} \qquad \text{or} \qquad q_c := \inf\{q : \bar{\mathcal{M}}(q) = 0\} \ . \tag{4.11}$$

It also corresponds to $\mathcal{M}(q) > 0 < \bar{\mathcal{M}}(q)$ for $q < q_c$, and $\mathcal{M}(q) = 0 = \bar{\mathcal{M}}(q)$ for $q > q_c$.

Let us investigate the distribution of clusters of the 4-regular periodic graph (4RP) introduced above for both site and bond percolation processes. To carry out the numerical analysis we basically proceed in the following way: First, we remove at random a certain fraction q_1 of vertices ($\bar{q}_1$ of edges) from our subject network. When we remove vertices in site percolation, all edges entering the removed vertices are also eliminated, that is, they are considered non occupied or blocked. However, when we remove edges in bond percolation, the vertices are not erased; note that the final network ($\bar{q} = 1$) will be an empty graph. Second, we count the number of vertices and the number of edges of all components of occupied vertices in which the network breaks up. The algorithm used to count vertices or edges is as follows: Once the fraction of vertices q_1 (of edges $\bar{q}_1$) is removed, a vertex from the remaining subnetwork of occupied vertices is selected. This vertex will act as root. Then, similar to the procedure applied when studying the tomography of networks, we follow the edges of the root to find all its neighbors (shell 1). Next, all edges of the first layer are followed until all vertices belonging to the second shell are found, etc. This process is executed until the cluster which the root belongs to is exhausted. The order and size of the cluster is obtained simply by counting the number of vertices and edges that are found during the process. The whole procedure is repeated by picking new roots (among those occupied vertices which are not yet counted in any cluster) until all components of the network are identified. By doing this, we can obtain the order and size of all components belonging to the subnetwork of non removed vertices. To obtain $\mathcal{M}$ and $\bar{\mathcal{M}}$ as a function of the removed fraction of vertices q (of edges $\bar{q}$), we take again the network from which the fraction q_1 was already eliminated and continue removing vertices until the next desired fraction q_2 is reached, and so on.

Figure 4.3 shows the behavior of the vertex and edge masses for some 4RP networks when varying the fraction q of removed vertices. We examine here 4RP networks of 10^4 vertices (green

[3]Whose both end-vertices are occupied vertices.

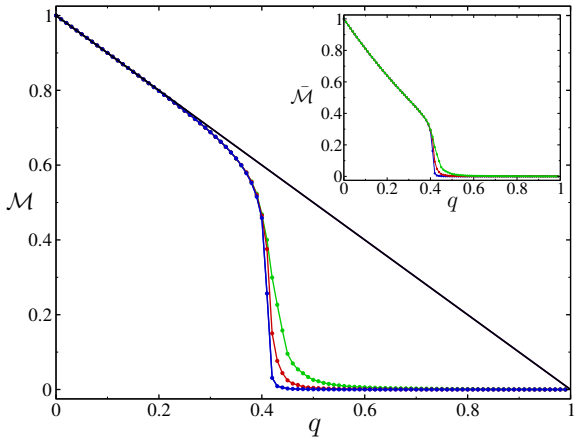

Figure 4.3: Site percolation. Large figure: Fraction of occupied vertices in the largest component, $\mathcal{M}$, versus the fraction of removed vertices q. The diagonal straight line corresponds to $\mathcal{M} = 1 - q$. Inset: Fraction of remaining edges as function of q. The networks considered are 4RP graphs. In both graphs: $N = 10^4$ (green curves), $N = 99856$ (red curves), and $N = 10^6$ (blue curves).

curves), $316^2 = 99856 \simeq 10^5$ vertices (red curves), and 10^6 vertices (blue curves). (Examining networks of different order is fundamental. By definition, a percolating component exists at a certain fraction q of removed vertices when their order is proportional to that of the entire network [Cohen et al., 2000a]. This, translated to infinite networks, means that vertex mass $\mathcal{M}$, and edge mass $\bar{\mathcal{M}}$, must be strictly positive for q; translated to finite networks, this means that $\mathcal{M}$ and $\bar{\mathcal{M}}$ must remain constant for the fixed fraction q of removed vertices when varying the order of the network N. Note that the definition of percolating component is a direct consequence of the assumption that the system is self-similar once the spanning component has appeared.) The simulation results have been obtained by averaging, respectively, over 20 ($N = 10^4$), 10 ($N \simeq 10^5$), and 5 ($N = 10^6$) realizations of the percolation process. In the figure, $\mathcal{M}$ (principal picture) and $\bar{\mathcal{M}}$ (inset) are plotted as function of q. The straight line in the principal picture represents the number of occupied (not yet removed) vertices of the network as function of q. We see that for small q all occupied vertices form only one component (which contains $N(1-q)$ vertices). As q increases, small clusters begin to separate from the largest component; however, the fraction of occupied vertices belonging to the largest component $\mathcal{M}(q)$ does not change as the order of the network increases. In fact, an inspection of the order distribution of clusters shows that this distribution remains constant irrespective of N. This behavior continues until q reaches the value $q \sim 0.40$, from whereon the curves begin to separate from each other (see also figure 4.5). Then, the fraction $\mathcal{M}$ falls with increasing speed as the order of the network grows. The inset shows the same qualitative behavior. Note that the point at which the system stops exhibiting the same $\bar{\mathcal{M}}$ is $q \sim 0.4$, as shown in figure 4.5.

We now examine the behavior of the preceding 4RP networks under bond percolation. We employ the same masses $\mathcal{M}(\bar{q})$ and $\bar{\mathcal{M}}(\bar{q})$ as above, but now as a function of $\bar{q}$, the fraction of removed edges. The critical fraction $\bar{q}_c$ can be obtained from the conditions $\bar{q}_c := \inf\{\bar{q} : \mathcal{M}(\bar{q}) = 0\}$ or $\bar{q}_c := \inf\{\bar{q} : \bar{\mathcal{M}}(\bar{q}) = 0\}$, the reason being, if a phase transition exists, it will satisfy $\mathcal{M}(\bar{q}) > 0 < \bar{\mathcal{M}}(\bar{q})$ for $\bar{q} < \bar{q}_c$, and $\mathcal{M}(\bar{q}) = 0 = \bar{\mathcal{M}}(\bar{q})$ for $\bar{q} > \bar{q}_c$. Figure 4.4 shows the vertex and edge masses $\mathcal{M}$ and $\bar{\mathcal{M}}$ as function of $\bar{q}$. As before, the results are averaged over 20 ($N = 10^4$), 10 ($N \sim 10^5$), and 5 ($N = 10^6$) independent realizations of the bond percolation process. We observe the same qualitative behavior as that corresponding to site percolation. In this case, however, the destruction of the largest component happens at $\bar{q} \sim 0.5$, and it is more distinguished than for site percolation. The entire subnetwork of occupied vertices consists of only one component until the fraction $\bar{q}$ of removed edges reaches the value $\bar{q} \sim 0.4$ -the straight line of the inset corresponds to $1 - \bar{q}$, the fraction of occupied or passable edges in the network. All three curves coincide for $\bar{q} < 0.5$, and at the point $\bar{q} = 0.5$, the largest component dwindles

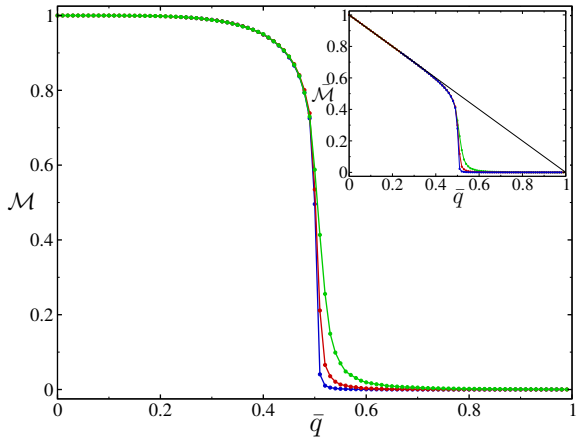

Figure 4.4: Bond percolation. Large figure: Fraction of occupied vertices in the largest cluster, $\mathcal{M}$, versus the fraction of removed edges $\bar{q}$. Inset: Fraction of passable edges as function of $\bar{q}$. The diagonal straight line corresponds to $\bar{\mathcal{M}} = 1 - \bar{q}$. Green curves: $N = 10^4$, red curves: $N = 99856$ vertices; blue curves: $N = 10^6$. The networks analyzed are the same as considered for site percolation.

in size as the order of the network grows (see figure 4.5); for larger $\bar{q}$ the system is no more than a set of many small clusters.

We can asses more explicitly the behavior for large fractions of removed vertices (q) and edges ($\bar{q}$) in figure 4.5. Both panels show $\mathcal{M}$ as a function of q (panel a) and $\bar{q}$ (panel b). (The logarithmic scale in the y-axis was introduced in order to observe more clearly the behavior for large q and $\bar{q}$, which, according to the classical percolation theory, must be exponential. We refrain from including the pictures corresponding to $\bar{\mathcal{M}}$ versus q and $\bar{q}$ since they show results similar to those exhibited in figure 4.5.) Let us interpret the results. This time, we describe the percolation behavior of networks as a function of the increasing p (as in lattice percolation): Considering that vertices of the network were initially non occupied ($p = 0$), and then, as p increases, progressively occupied at random until all of them were occupied ($p = 1$). Hence, we proceed as if we read the pictures from right to left (since $q = 1-p$). Thus, we observe that, as p begins to increase from $p = 0$ on, the order of the largest component grows in an approximately exponential way, both for site and bond percolation (notice the logarithmic scale). This behavior shows excellent agreement with the heuristic percolation law affirming that large components grow exponentially from $p = 0$ on, as long as p is still far way from p_c [Stauffer and Aharony, 1994]. Then, as the fraction of occupied vertices continues to increase, a moment comes in which the largest component stops growing exponentially but yet turns to more accelerated growth until, at certain critical p, it percolates (to the effect that $\mathcal{M}$ and $\bar{\mathcal{M}}$ are proportional to the order of the network N). This intensive growth becomes more sudden the large the order of the network. From the critical p on, the already percolating component continues growing rapidly[4], but not as excessively as when on the verge of percolating. For larger p, the spanning component grows almost linearly with p. Thus, two different regions can be clearly distinguished in the pictures: a region in which all plotted curves coincide, which evidences the existence of the spanning component in this region, and another region where the curves are largely separated, in which masses $\mathcal{M}$ and $\bar{\mathcal{M}}$ tend to zero as the order of the network increases. When extrapolating this behavior to infinite networks ($N \to \infty$), the tendency indicates the existence of a true percolation threshold, q_c ($\bar{q}_c$), close to the point at which the curves begin to separate. Thus, for $q < q_c$ ($\bar{q} < \bar{q}_c$), a spanning component containing a non-zero fraction of the infinite vertices of the network exists, while for $p > p_c$ ($\bar{q} > \bar{q}_c$), no percolating cluster exists, and all components of the system have a finite number of elements.

This conclusion is also supported by the scaling behavior close to the critical q_c ($\bar{q}_c$). In

[4]In fact, approximately following a power law, as we will later see.

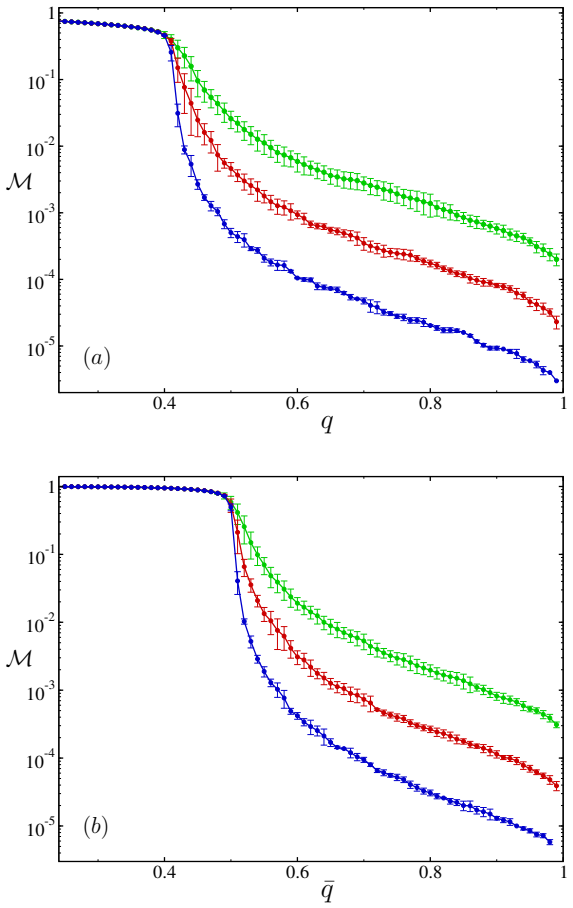

Figure 4.5: (a): Site percolation. Fraction of occupied vertices $\mathcal{M}$ as function of q. (b): Bond percolation. Fraction of occupied vertices $\mathcal{M}$ as function of $\bar{q}$. Note the logarithmic scale in both pictures. Green curves correspond again to $N = 10^4$, red curves to $N \sim 10^5$, and blue curves to $N = 10^6$. In both graphs there is a point which separates two clearly distinguishable regions. Below this critical point $\mathcal{M}$ and $\bar{\mathcal{M}}$ do not vary when the order of the network changes; above it, both tend to zero as $N \to \infty$ (see text for more details).

lattice percolation, the spanning cluster grows as power law like $P_\infty \sim (p - p_c)^\beta$, when $p \to p_c^+$ (see Eq. 4.5). This scaling law can be rewritten in terms of q as

$$\mathcal{M} \sim (q_c - q)^\beta \quad , \quad q \to q_c^- \qquad \text{and} \qquad \bar{\mathcal{M}} \sim (q_c - q)^{\beta'} \quad , \quad q \to q_c^- \ , \tag{4.12}$$

equations which only state that the fractions of non removed vertices, $\mathcal{M}$, and edges, $\bar{\mathcal{M}}$, in the largest component must drop to zero as power laws. Equivalent equations for bond percolation can be written changing $q \to \bar{q}$. Figure 4.6 shows the fits of $\mathcal{M}$ (green) and $\bar{\mathcal{M}}$ (yellow) as function of $0.4085 - q$ and $0.5005 - \bar{q}$ respectively, for those intervals of q (site percolation) and $\bar{q}$ (bond percolation) in which $q \to q_c^-$ and $\bar{q} \to \bar{q}_c^-$. Both fits were accomplished by using the data corresponding to the network of 10^6 vertices. Similar results are obtained using data of the other networks of $N = 10^4$ and $N \sim 10^5$. We see that, on double logarithmic scales, the simulated points follow a straight line, which confirms the expected power law decay. The fits provide the critical thresholds $q_c = 0.4085$ and $\bar{q}_c = 0.5005$, and the critical exponents $\beta = \beta' \simeq 0.15$.

These critical thresholds are merely good approximations of the real concentrations q_c and $\bar{q}_c$, at which the phase transitions really take place. The numerical simulations (see figures 4.3, 4.4 and 4.5) show that the (finite) networks considered do not undergo sharply defined percolation thresholds. A more exact determination of the critical concentrations requires the previously described real-space renormalization techniques or the finite-size scaling formalism. Unfortunately, both approaches involve certain difficulties which are not easy to overcome in the more general case of networks with complex topologies. Thus, for applying the renormalization procedure, it must be possible to carry out the above mentioned "blocking" process, that is, to

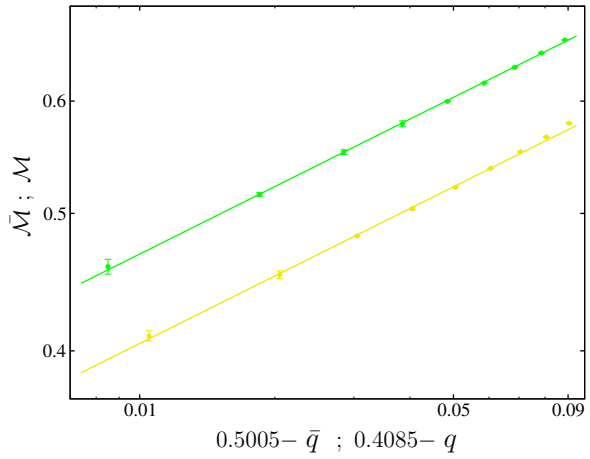

Figure 4.6: Power law decay of the vertex and edge masses close to the percolation thresholds q_c and $\bar{q}_c$, respectively. Upper curve: Site percolation; $\mathcal{M}$ *versus* $0.4085 - q$. Lower curve: Bond percolation; $\bar{\mathcal{M}}$ *versus* $0.5005 - \bar{q}$. Note the double logarithmic scales.

divide the vertices of the network into blocks so that the resulting network be like the original ones. This is, in general, quite a difficult task due to the non-regular vertex degree distribution and the presence of diverse types of correlations. On the other hand, the finite-size scaling formalism does not give accurate outcomes in this case. In order to successfully apply this formalism the effective thresholds q_{ef} must be precisely calculated. However, because in finite networks the percolation transitions are not sharp enough, this is not simple. The principal goal of this study, however, is to determine if certain complex networks undergo phase transitions or not, and not so much to calculate exactly when they take place. With respect to this, the above analysis demonstrates without doubt the existence of site and bond percolation thresholds for the 4RP networks. Note, however, that our results $q_c = 0.4085$ and $\bar{q}_c = 0.5005$ are not bad: the currently accepted thresholds for the square lattice (to which our results must tend when $N \to \infty$) are $q_c^{sl} = 0.40725$ and $\bar{q}_c^{sl} = 0.5$.

Let us, however, commence with the statistical study of the largest cluster. We concluded that $\mathcal{M}$ as well as $\bar{\mathcal{M}}$ vanish for the first time at the same critical point, q_c, when $N \to \infty$. We ask whether this is the only remarkable property of the system near q_c. Figure 4.7 compares the order and size of the largest component as a function of the fraction q of removed vertices (panel *a*) and the fraction $\bar{q}$ of removed edges (panel *b*) during the percolation process. The insets show the same results on a logarithmic scale on the y-axis, so that the behavior of $\mathcal{M}$ and $\bar{\mathcal{M}}$ for large values of q ($\bar{q}$) are clearly visible. In the pictures, orange curves reproduce the behavior of the vertex mass, $\mathcal{M}$, and cyan curves[5] the fraction between the number of edges in the largest component and N. We can observe again two regions in the pictures showing different qualitative behaviors: For $q < q_c$ ($\bar{q} < \bar{q}_c$), the number of edges is visibly larger than the number of vertices. Both quantities, the order and the size of the largest component, tend to be equal as q ($\bar{q}$) approach q_c ($\bar{q}_c$); the ratio between both quantities decreases from the value $\langle k \rangle$ at $q = 0$ ($\bar{q} = 0$) to approximately one at the critical point q_c ($\bar{q}_c$). For $q > q_c$ ($\bar{q} > \bar{q}_c$), the largest component[6] is approximately a tree. (Some simulations show the existence of one or two cycles in the largest cluster above the threshold, but very probably the number of cycles tends to zero when $N \to \infty$). This latter can be more clearly observed in the insets, where cyan and orange curves coincide for all q ($\bar{q}$). All these results correspond to the networks of 10^6 vertices; the same behavior, however, can be observed for the networks of 10^4 and 10^5 vertices. We will

[5]Thus, cyan curves do not really show the behavior of $\bar{\mathcal{M}}$, as is marked in the pictures, but the behavior of a quantity proportional to it. We refrain from introducing a new symbol for this quotient in order not to annoy the reader with too many definitions.

[6]Also all other clusters of the network

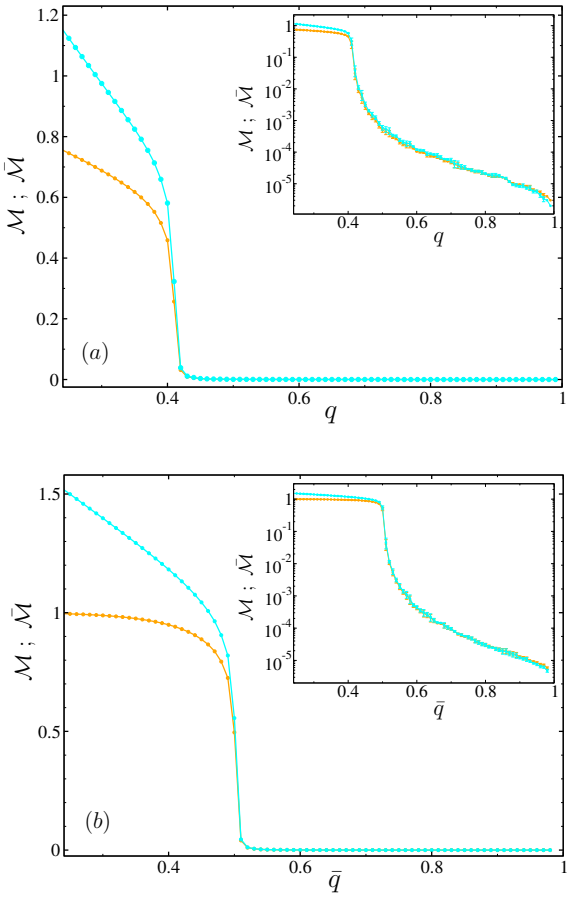

Figure 4.7: All cyan curves are proportional to the number of edges in the largest cluster; the orange ones are proportional (with the same constant of proportionality) to the number of vertices in the largest cluster (see text). Thus, the four pictures compare the behavior of the order of the largest cluster with its size. Panel (a): Site percolation. Panel (b): Bond percolation. The insets show the same as the corresponding graphs, but on a logarithmic scale on the y-axis. We see in all pictures the existence of a special point which separates two qualitatively different behaviors: below the point the size is always larger than the order; above it the largest cluster is approximately a tree. These points coincide with the percolation thresholds we previously found.

see that this characteristic is exhibited by all networks we investigate.

The explanation why this happens is intuitively simple. Let us explain it in terms of the fraction of occupied vertices p (instead of q). We then assume that, at $p = 0$, all vertices of the network are empty, and, as p increases, vertices are progressively occupied at random. (We discuss here the case of site percolation, however, after making suitable changes, the argument is also valid for bond percolation). When the fraction of occupied vertices p is very small, the network consists of many isolated occupied vertices which are distributed at random in the system; these occupied vertices are completely surrounded by non-occupied neighbors. As p increases, the probability that one of the neighbors is occupied grows, which gradually produces small components composed of occupied vertices. The order of a component grows if and only if an empty neighbor of the component is occupied. Assume that this occurs. In the case in which p is small, the order and size of the component will typically increase only by unity, as almost all neighbors of the empty selected vertex will be non-occupied vertices, and, usually, only one of them will be occupied. (Bear in mind that no edge can be occupied if its two end-vertices are occupied.) Thus, for p small, the emerging components will typically present a structure similar to trees. In this actual state of the system, the distance between any two components is quite large, and consequently, each component grows proportionally to the number of non-occupied vertices which surround it. If components are not too small, it can be proven that they must grow in an approximately exponential way [Stauffer and Aharony, 1994]. The situation changes near the critical concentration p_c. Here, the fraction of occupied vertices is already large enough so that most components are separated by only a few non-occupied vertices. When these vertices

are occupied, they typically act as bridges joining the different components. This way, the components of the network, and particularly the largest component, undergo a sudden growth when p only slightly increases. This intensive growth stops when the percolating cluster appears. At this moment, the appearance of the system is exactly opposite to p small: now, occupied vertices begin to surround non-occupied vertices. Thus, when a new vertex is occupied, it introduces a quite large number of edges to the system, due to the large of its occupied neighbors. Thus, around the same time at which the percolating component appears, the first cycles begin to appear. (For p smaller, the fraction of circles in the network is negligible.) As $N \to \infty$, the region where the largest component suddenly begins to grow, diminishes until it tends to concentrate on only one point: the percolation threshold p_c. (Note that the cycles occur only if the initial network is not a tree). As we continue to occupy more and more vertices, the occupied vertices begin to almost completely surround the non-occupied vertices; at this moment, the order of the largest component grows almost linearly with p, and almost all edges of any newly occupied vertex become occupied.

Random graphs. Before we investigate the percolation properties of scale-free networks, we describe briefly the behavior of phase transitions in random graphs, the only type of networks for which some analytical solutions related to percolation exist. Remember that, to construct a random graph, we start from N isolated vertices and then begin randomly adding edges to the system. Following the results of Erdös and Rényi, we expect the appearance of the giant component when the number of introduced edges, L, reaches the value $L = N/2$, and a graph totally connected at $L = N\ln(N)/2$. For our study, we consider three initially empty graphs, of 10^4, 10^5, and 10^6 vertices respectively, and then we add edges at random (avoiding loops and multiedges) until the number of edges is five times larger than the number of vertices in the corresponding network. We then expect to find a phase transition at $\bar{q}_c = 1 - N/(2L) = 0.9$, and a fully connected network at $\bar{q}_c = 1 - N\ln(N)/(2L)$. When translating this into our present language, these results are evidently related to the bond percolation. There are no theoretical studies dealing with site percolation on random graphs. We can only affirm that a phase transition at certain $q_c \leq \bar{q}_c$ must exist [Kesten, 1982].

We show in figure 4.8 how the masses of the largest component $\mathcal{M}$ and $\bar{\mathcal{M}}$ behave as function of the removed fractions of vertices, q, and edges, $\bar{q}$ (for site and bond percolation, respectively). The pictures show that our simulation results are consistent with the theory. Thus, panel (b) evidences a sudden change in the order and size of the largest cluster close to $\bar{q} = 0.9$; on the other hand, panel (d) shows how $\mathcal{M}$ rapidly tends to zero as $N \to \infty$. The same behavior can be observed for $q = 0.9$ for site percolation in the upper panels (a) and (c). All results are averaged over 20 ($N = 10^4$), 10 ($N = 10^5$), and 5 ($N = 10^6$) realizations of the corresponding percolation processes. Although we do not show the corresponding pictures, the same qualitative behavior can be found when analyzing the fraction of edges $\bar{\mathcal{M}}$ existing in the largest cluster. We also fitted the decay of $\mathcal{M}$ and $\bar{\mathcal{M}}$ for both types of percolation to power law functions. We find that, as well as the order, also the size of the largest cluster follows a power law for all N investigated. Interestingly, the fits show a small shift of the effective threshold q_{ef} to the positive direction of q, from $q_{ef} \simeq 0.889$ (for the graph of 10^4 vertices) to $q_{ef} \simeq 0.898$ (for the graph of $N = 10^6$). Note, however, that the shift of the effective threshold is very small (≤ 0.01), and that q_{ef} always remains below the percolation threshold $\bar{q}_c = 0.9$. For bond percolation, decay is so fast that $\mathcal{M}$ seems to fall more rapidly than a power law; however, the existence of the phase transition is in this case not disputed since this was analytically proven by Erdös and Rényi.

Theoretical studies about random graphs not only conclude the existence of the phase transition, but also, before the spanning cluster appears in the system, that the network is only composed of isolated trees (Albert and Barabási [2002a]). This aspect is confirmed by our sim-

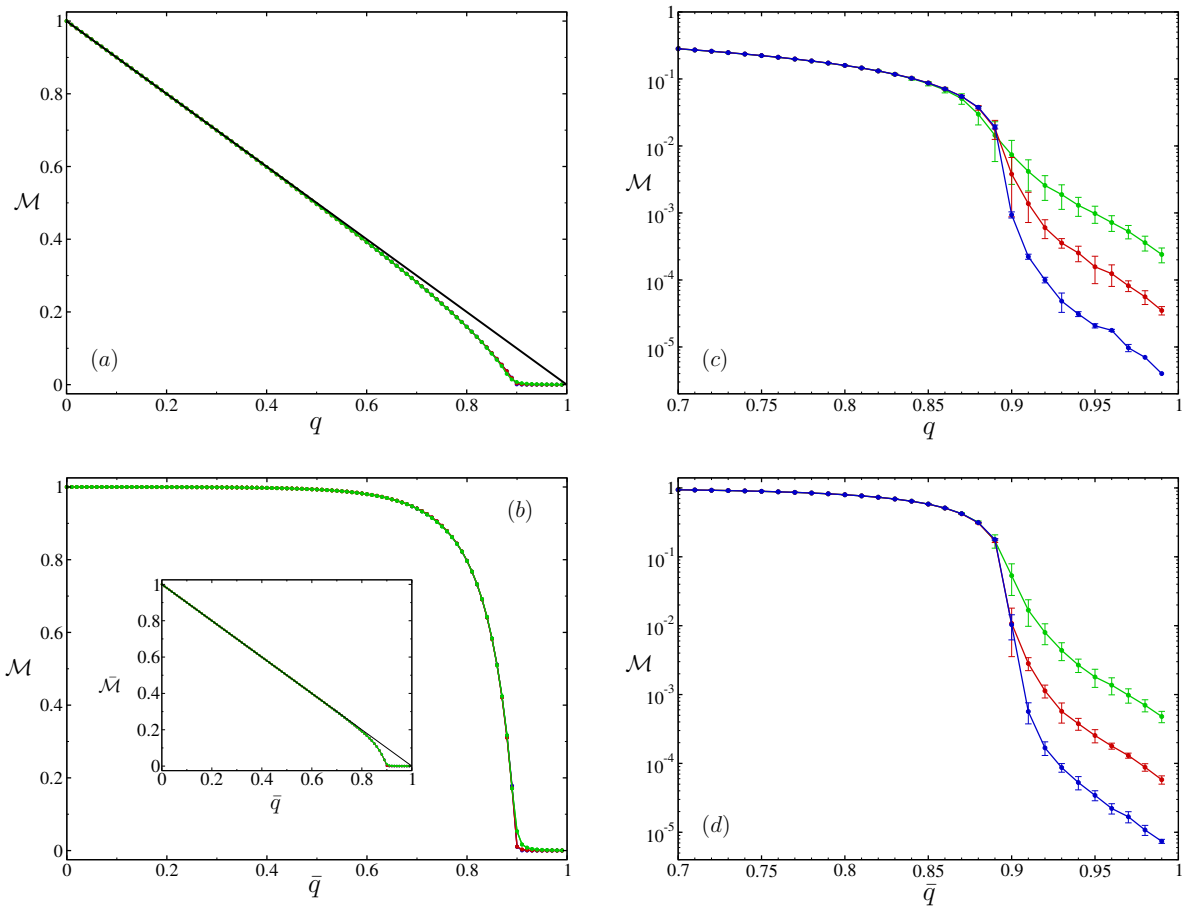

Figure 4.8: (a): Site percolation. $\mathcal{M}$ as function of q. (b): Bond percolation. $\mathcal{M}$ *versus* $\bar{q}$. In the inset is shown $\bar{\mathcal{M}}$ as function of $\bar{q}$. (c): Site percolation. $\mathcal{M}$ *versus* q on a logarithmic scale on the y-axis. (d): Bond percolation. $\mathcal{M}$ *versus* $\bar{q}$ on a logarithmic scale on the y-axis. All pictures contain three curves, corresponding to the results of random graphs: $N = 10^4$ (green), $N = 10^5$ (red), $N = 10^6$ (blue).

ulations. All examined graphs, the small ones with $N = 10^4$ as well as the largest graphs of 10^6 vertices, show that random graphs are *forests* above $\bar{q}_c = 0.9$. The same occurs in the case of site percolation, where we cannot find a single cycle beyond $q = 0.89$. Thus, we observe the same behavior we already found when studying the 4RP graph: the fraction of cycles in the system is infinitesimal above the percolation threshold. Inspired by these results, we define a new measure for detection when the largest cluster passes from a component with cycles to a tree; since the crossing usually happens at the percolation threshold, this measure can also be used to characterize the phase transition. The quantity -which is not a probability- is

$$\mathcal{D} = \frac{\text{(number of edges - number of vertices) in the largest cluster}}{\text{number of vertices in the entire network}} \quad , \tag{4.13}$$

where the only purpose of the denominator is to in some way normalize the results of networks of different order, so that they can be compared to each other. In figure 4.9, we plot the difference D as function of q (panel a), and $\bar{q}$ (panel b) for the random graph with 10^6 vertices. The same results are obtained for other networks with $N = 10^4$ and $N = 10^5$. In both cases the largest cluster is a tree above the respective percolation thresholds q_c and $\bar{q}_c$. Moreover, a fit of D as

function of $0.9 - \bar{q}$ (bond percolation), and $0.9 - q$ (site percolation) shows that the decay of D follows a power law. The insets of the pictures show, as before, the number of edges divided by the number of vertices of the largest cluster (cyan curves) and $\mathcal{M}$ (orange curves), in order to compare them near the threshold.

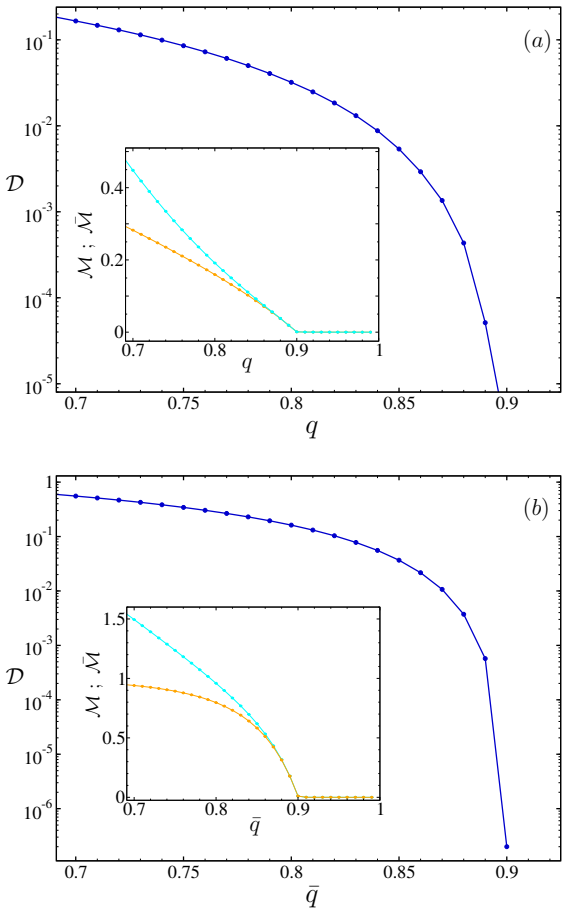

Figure 4.9: Large pictures: $\mathcal{D}$ as function of q (panel a, site percolation), and $\bar{q}$ (panel b, bond percolation). The pictures show that above the corresponding critical thresholds the largest component is a tree (see text). The insets show how the number of edges (cyan) and vertices (orange) of the largest component behaves with q (panel a, site percolation), and $\bar{q}$ (panel b, bond percolation). Both curves are normalized by the number of vertices in the network. (see also fig. 4.7). All results correspond to a random graph of 10^6 vertices.

The fact that below the critical threshold the largest cluster contains cycles, while above the threshold it is a tree, can be used to generalize the above introduced measure S, the *mean cluster size*. We define the new parameter $\mathcal{S}$, akin to the mean cluster size, as:

$$\mathcal{S} = \sum_s s w_s = \frac{\sum_s s^2 n_s}{\sum_s s n_s} \ . \tag{4.14}$$

where n_s is again the number of s-cluster per unit site. Now, however, the sum must run over all clusters of the system except for the largest component if it is not a tree. Thus, below the percolation threshold q_c the sum will run over all clusters except for the largest one, and above it, over all clusters. Note that this constraint is equivalent to restricting the sum to the finite clusters when $N \to \infty$. A plot of this new quantity, $\mathcal{S}$, versus the fraction of removed vertices must then show (as occurred for lattices) a peak at q_{ef} even for finite networks. This can be used as an additional indication of the existence -or non-existence- of the phase transition. (In fact, a measure of $\mathcal{S}$ versus q ($\bar{q}$) for our previous 4RP networks exhibits a pronounced peak at q_c ($\bar{q}_c$), despite the fact that in this case clusters are only "approximately trees" beyond the threshold.)

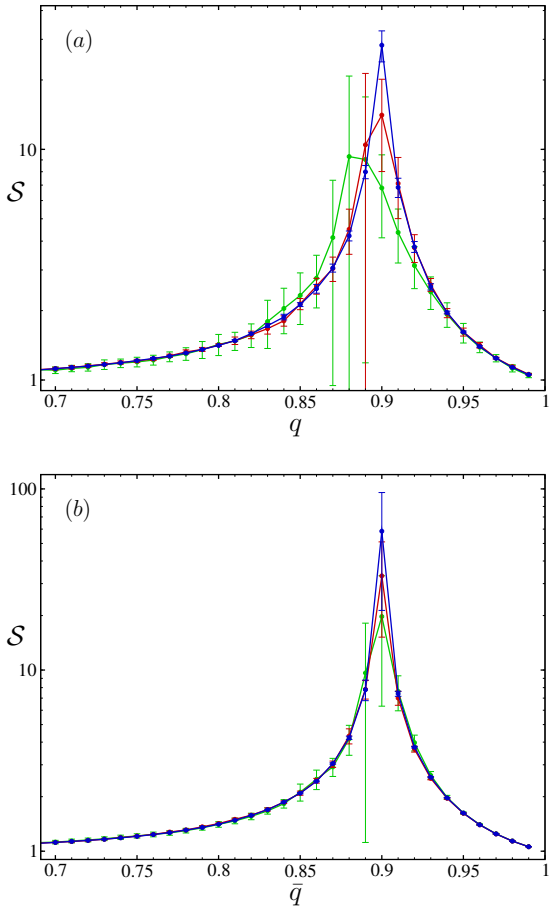

Figure 4.10: Panel (a): Site percolation. Generalized mean cluster size $\mathcal{S}$ as a function of the removed fraction of vertices q. Panel (b): Bond percolation. Generalized mean cluster size $\mathcal{S}$ as a function of the removed fraction of edges $\bar{q}$. The two pictures contain the three typical curves: green ($N = 10^4$), red ($N = 10^5$), and blue ($N = 10^6$). For bond as well as site percolation the results show a peak close to the percolation points $q_c \sim 0.9$ and $\bar{q}_c = 0.9$. Note that the peaks move towards the growing q (site percolation) as the order of the network increases; this shift is, however, almost negligible for large random graphs. For bond percolation all curves peak at $\bar{q}_c = 0.9$.

The next figure 4.10 shows the behavior of $\mathcal{S}$ near the critical threshold as function of q (panel a), respectively $\bar{q}$ (panel b). As before, green curves correspond to results of networks with $N = 10^4$, red curves to $N = 10^5$, and blue ones to $N = 10^6$. In the upper panel, corresponding to site percolation, we can see the existence of a small shift of the peaks towards growing q, in agreement to our results obtained from the power law decay of $\mathcal{M}$. This shift is, however, only noticeable for small networks; for large random graphs, like the ones with $N = 10^5$ and $N = 10^6$, the shift is negligible, which can be interpreted as if the effective thresholds, q_{ef}, already lie very close to the percolation threshold, q_c, corresponding to the thermodynamic limit. This shift is still more negligible in bond percolation: we see in the lower panel that all curves, even the one corresponding to the smallest network ($N = 10^4$) peaks at $\bar{q}_c = 0.9$). This effect will also later be observed in scale-free networks. (Thus, since it is always verified that $q_c < \bar{q}_c$, we can even investigate the behavior of the bond percolation case even for deciding the existence of phase transitions in site percolation processes. The study of the bond percolation is usually more precise than the analysis of the system under site percolation, given that the results of bond percolation present smaller biases due to finite-size-effects.)

4.4 Percolation on scale-free networks

The robustness of uncorrelated scale-free networks has been studied in the past few years by several groups [Cohen et al., 2000a, 2002a, Albert et al., 2000, Cohen et al., 2000b, Schwartz

et al., 2002]. They investigate the vulnerability of networks to targeted attacks and the random breakdown of vertices. Most of them, however, are heuristic studies not sufficiently supported by simulations. One of the aims of this section is to corroborate the results of these studies of uncorrelated networks. The principal goal, however, is to assess the influence of vertex-pair correlations on the percolation properties [Vázquez and Moreno, 2003, Xulvi-Brunet et al., 2003, Xulvi-Brunet and Sokolov, 2004]. Extensive numerical simulations indicate that correlations play an essential role in percolation processes, notably affecting the critical point at which the percolation transition takes place. Other interesting studies about percolation on geographical scale-free networks can be found in [Huang et al., 2005a, Sander et al., 2002].

In order to study the percolation of the vertex-pair correlated networks, we consider some network models to which the algorithms AS, DS, and TS are applied. The network models we take into consideration are the following: Model 1, the *intrinsic-worth* model. This is the modified Barabási-Albert construction with initial attractiveness, whose degree distribution tends to $P(k) \sim k^{-3-A/m}$ (see discussion in chapter 2). In this case, we take $A = 4$ and $m = 2$ (as in the example we presented in the second chapter) so that the second moment $\langle k^2 \rangle$ of the degree distribution is the finite value $\langle k^2 \rangle = 28$; Model 2, the *exact scale-free* construction. In this the model each vertex i is assigned an entire k_i drawn from the desired probability distribution $P(k) = ck^{\gamma}$ (see again in chapter 2). We choose $\gamma = 3.5$, so that $\langle k^2 \rangle$ is again finite; Model 3, the current Barabási-Albert construction. Here, the networks are generated using $m = 2$. Remember that the Barabási-Albert model exhibits $\langle k^2 \rangle \to \infty$ as $N \to \infty$; Model 4, the *exact scale-free* construction again, but in this case taking $\gamma = 2.5$ in order to construct exact scale-free networks with diverging second moment.

Randomized scale-free networks. Cohen et al. [2000a, 2002a] argue that, for uncorrelated networks, the critical breakdown threshold may be found by the following criterion: "if circuits of connected vertices may be neglected, the percolation transition takes place when a vertex, i, connected to a vertex j in the spanning cluster, is also connected to at least one other vertex; otherwise the spanning cluster is fragmented". The authors conclude then that the critical concentration q_c verifies

$$1 - q_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \ , \tag{4.15}$$

where $\langle k \rangle$ and $\langle k^2 \rangle$ are, respectively, the first and second moment of the degree distribution of the original network before the random breakdown. The authors consider random breakdowns of vertices (but not of edges), so that the condition only corresponds to site percolation. Consequently, uncorrelated networks exhibiting a diverging second moment, like the scale-free networks whose exponent satisfies $\gamma \leq 3$, have a percolation threshold $q_c = 1$ in the thermodynamic limit. In epidemiological terms, this corresponds to the absence of herd immunities in such systems.

Figure 4.11 shows the percolation behavior of uncorrelated versions of the four models considered. Random correlations are generated on these networks by applying the RS algorithm. The curves show the behavior of $\mathcal{M}$ for (from bottom to top) the model 2 -exact scale-free network with finite second moment, $\gamma = 3.5$-, model 1 -evolving network with attractiveness-, model 3 -the original Barabási-Albert construction-, and model 4 -exact scale-free network with infinite second moment-. All results correspond to networks having $N = 10^5$ vertices, and are averaged over ten independent realizations of the RS algorithm. Panel (a) represents the results corresponding to the site percolation, and panel (b), those corresponding to the bond percolation; the inset is an enlargement of the inferior part on the right of panel a. Green and red curves (models 1 and 2) undergo, both for site and bond percolation, an abrupt change in their

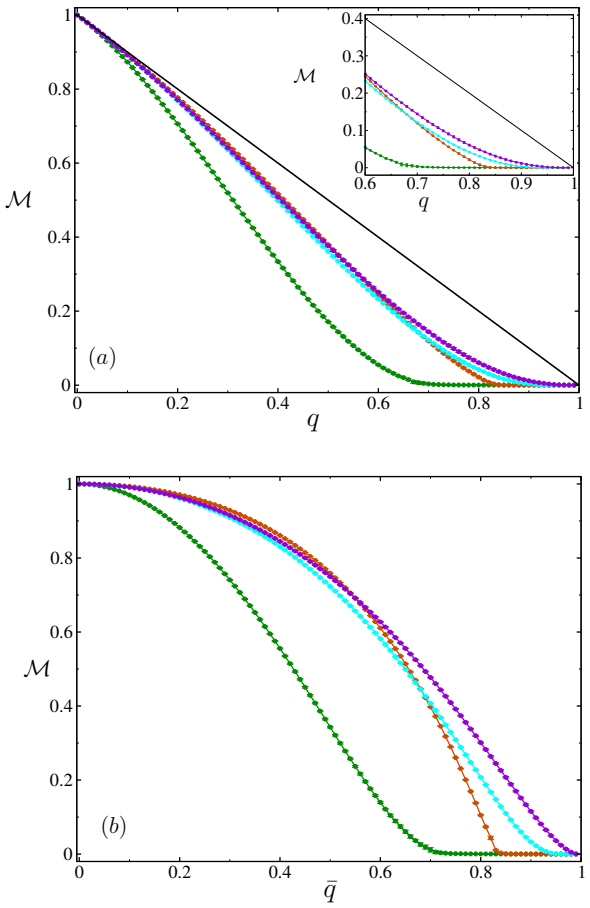

Figure 4.11: Percolation behavior of uncorrelated scale-free networks. (a): Site percolation. The inset is an enlargement of the right inferior part of panel a. (b): Bond percolation. In both percolation cases the curves represent the behavior of the following network models: model 1 (dark green), model 2 (red), model 3 (turquoise), model 4 (indigo). See text for more details about the models. The networks considered for the simulations have 10^5 vertices.

behavior close to $q \simeq 0.7 \simeq \bar{q}$ and $q \simeq 0.833 \simeq \bar{q}$, respectively. In the next subsection, where we analyze these results in more depth, and compare them with the corresponding correlated versions, (see figures 4.17 and 4.18), we will see that the points at which the curves undergo this change practically coincide with the points where the percolation transition should theoretically occur[7]. The upper curves of the pictures, however, do not show any sharp change. Therefore, we may not conclude from these data anything about the percolation thresholds of models 3 and 4.

The pictures evidence the following interesting percolation aspect of the scale-free networks: the larger the number of hubs the more robust the network is. Thus, model 4 contains more vertices of large degree than the Barabási-Albert construction (model 3), and this, in turn, has many more hubs than model 1 (the Babarási-Albert model with initial attractiveness); Model 2, the exact scale-free model whose degree distribution follows $P(k) \propto k^{-3.5}$ (finite second moment), contains very few hubs; The existence of a large amount of highly connected vertices in the networks 3 and 4 explains the smooth behavior of $\mathcal{M}$, which we find in the curves of figure 4.11. Sure enough, as more and more vertices of the network become occupied, the probability that separated small components joined to one large component is large, since hubs operate as bridges between the small components. As a result of that, relatively large components appear

[7]Note that this abrupt change in the behavior of $\mathcal{M}$ is sharper for the curves corresponding to bond percolation. This fact, which can be observed in all our simulation results, may be exploited for i) determining (a first estimation of) the bond percolation threshold $\bar{q}_c$, and ii) setting an upper limit of the critical threshold q_c in the case of site percolation, in which the change is usually not so evident.

in the system even for small fractions of p, and, when occupying more and more vertices, these components rapidly fuse into a giant component which dominates the system. However, as we will see, vertex-pair correlations change the features of the problem considerably: the critical fraction at which the spanning cluster appears for the first time depends on the neighbors of the hubs.

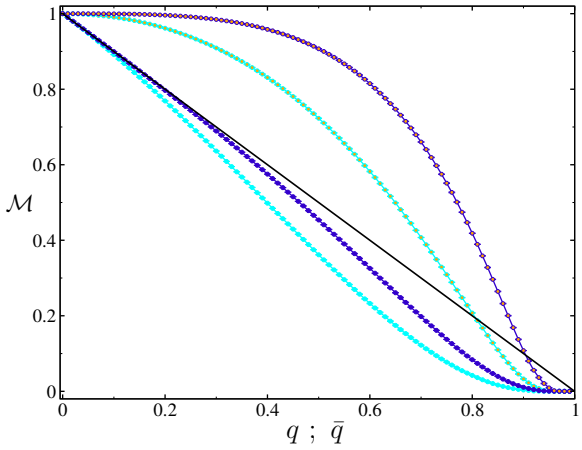

Figure 4.12: Fraction of vertices in the largest component $\mathcal{M}$ as a function of: the removed fraction of vertices (the two upper curves of the picture), and the removed fraction of edges (the two lower curves). The networks are uncorrelated Barabási-Albert construction of $N = 10^5$, generated using $m = 2$ (cyan curves) and $m = 3$ (blue curves).

Another aspect which evidently influences the nature of percolation processes is the average number of connections in the network: the larger $\langle k \rangle$ is, the more robust the network is. This point is shown in figure 4.12. In this picture, the fraction of remaining vertices in the largest cluster $\mathcal{M}$ is plotted as a function of $\bar{q}$ (the two upper curves, corresponding to site percolation) and q (the two lower curves, corresponding to bond percolation). The networks we examine here are Barabási-Albert networks (model 3), generated using the values $m = 2$, corresponding to $\langle k \rangle = 4$ (cyan curves), and $m = 3$, corresponding to $\langle k \rangle = 6$ (blue curves); the order of the networks is $N = 10^6$, and all results are obtained averaging over ten realizations of the RS algorithm. We see that a relatively small increase in $\langle k \rangle$ produces a considerable increase in $\mathcal{M}$, both for site and bond percolation: the network becomes more robust. The fact is so obvious that it does not need to be explained; it is sufficient to bear in mind the extreme case in which the network is a complete graph: in this case $\mathcal{M}$ follows exactly the black diagonal line of the picture, and, since there is always only one component in the system containing $(1 - q)N$ (or $(1 - \bar{q})N$) vertices, no percolation threshold ($q = \bar{q} = 1$) exists.

Let us now examine the behavior of the networks for which $\langle k^2 \rangle$ diverges when $N \to \infty$. For this purpose, we take into consideration the preceding bond percolation results of the Barabási-Albert network generated using $m = 3$. The motivation to do this is twofold: On the one hand, figure 4.12 shows that the curve that undergoes sharper change in the behavior of $\mathcal{M}$ is precisely the curve corresponding to the bond percolation case and $m = 3$; we expect that in this case the percolation threshold is easier to detect. On the other hand, if a percolation threshold exists, then it also must exist for the other cases considered: $m = 2$ and/or site percolation. Figure 4.13 shows the findings of this analysis. The results correspond to networks with $N = 10^4$ (grey), $N = 10^5$ (magenta), and $N = 10^6$ (blue), and they are all averaged over ten realizations except for those networks with $N = 10^4$ averaged over 20 realizations. In the principal picture $\mathcal{M}$ is plotted as a function of $\bar{q}$. It shows the typical situation we have already observed in previous systems which exhibit a phase transition: there are two different regions in the picture, one in which all three curves coincide (the large interval which extends from $\bar{q} = 0$ to $\bar{q} \simeq 0.95$) and another region where $\mathcal{M}$ seems to tend to zero as $N \to \infty$ (from $\bar{q} \simeq 0.95$ to $\bar{q} = 1$). Moreover, the scaling formalism seems to confirm the existence of this true percolation threshold. If we

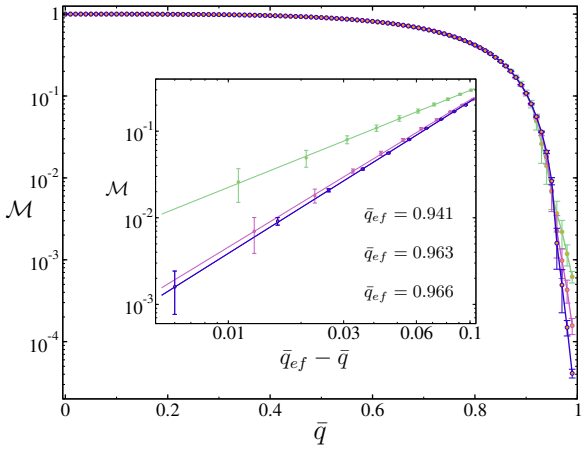

Figure 4.13: $\mathcal{M}$ as a function of $\bar{q}$. Uncorrelated Barabási-Albert networks generated using $m = 3$: $N = 10^4$ (grey), $N = 10^5$ (magenta), $N = 10^6$ (blue). Inset: Scaling analysis of the networks considered. Note that $\mathcal{M}$ decays as a power law when $\bar{q} \to \bar{q}_{ef}$.

take the values of $\mathcal{M}$ lying in the interval $q \to \simeq 0.95^{-}$ into account, and fit them using power law functions, then we obtain the corresponding "effective" percolation thresholds $\bar{q}_{ef}$. These fits are plotted on double logarithmic scales in the inset of the figure. The effective thresholds we obtain are: $\bar{q}_{ef} = 0.941$ for $N = 10^4$, $\bar{q}_{ef} = 0.963$ for $N = 10^5$, $\bar{q}_{ef} = 0.966$ for $N = 10^6$. By making use of the Eq. 4.6, it is now not difficult to calculate the value of the final percolation threshold: $\bar{q}_c \simeq 0.9667$.

Unfortunately, the values $\bar{q}_{ef}$ which we obtain by means of the fits suffer from a certain lack of precision. The inaccuracy is definitely slight, but large enough to make the scaling prediction of the Eq. 4.6 uncertain. Although our data indicates the existence of a real transition at $\bar{q}_c \simeq 0.97$, we cannot thus discard that the critical concentration is $\bar{q}_c = 1$. (The question could possibly be elucidated by larger simulations, using, for example, networks with $N = 10^7$ or $N = 10^8$ vertices, but nowadays computation time for such systems is prohibitive.) Thus, the important question on the existence of phase transitions in networks whose $\langle k^2 \rangle$ diverges must be studied in more depth in the future, both analytically and numerically.

On the other hand, figure 4.14 shows how the quantity $\mathcal{D}$ behaves as a function of $\bar{q}$. One can observe again that the largest component becomes a tree close to the obtained percolation threshold. Moreover, the simulations shows that the system has no cycles above $\bar{q} \simeq 0.95$. Note that the curves corresponding to the networks having $N = 10^5$ (magenta) and $N = 10^6$ (blue) practically coincide. This indicates that the system is already very close to the thermodynamic limit when $N = 10^6$, and supports the validity of our previous finding for $\bar{q}_c$. In the inset, $\mathcal{S}$ is plotted versus $\bar{q}$. Quantity $\mathcal{S}$ assumes that the largest component is always a tree above the percolation threshold $\bar{q}_c$, and that it contains a non-negligible fraction of cycles below $\bar{q}_c$ (if the network is not a forest). This assumption could of course be a double-dealer for diverging uncorrelated scale-free networks, in spite of the numerical evidence which is present for other types of networks. Therefore, the results of the inset may not be conclusive when it comes to assessing the existence of the phase transition. Nevertheless, it is quite interesting that $\mathcal{S}$ peaks close to $\bar{q} \simeq 0.95$ (for the system with $N = 10^6$), where the corresponding $\bar{q}_{ef}$ is located.

Exact scale-free networks with a diverging second moment could also be examined to determine the existence or inexistence of true percolation transitions in networks where $\langle k^2 \rangle \to \infty$. However, in our opinion, they are not networks suitable for investigating the problem. For this type of networks the first moment of the degree distribution $\langle k \rangle$ does not remain constant, but it grows as N increases (see chapter 2); this increase of $\langle k \rangle$ already produces an increment of the robustness of the network, which could sufficiently influence the analysis thus leading us to wrong conclusions about the role of the diverging second moment $\langle k^2 \rangle$. (Note that model 1

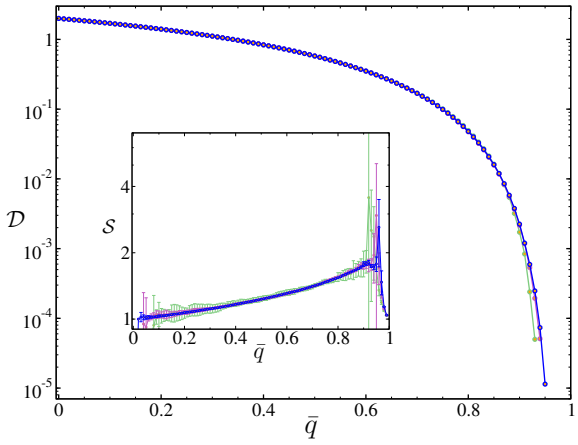

Figure 4.14: $\mathcal{D}$ (large graph) and $\mathcal{S}$ (inset) as a function of the removed edge fraction $\bar{q}$ (bond percolation) for the uncorrelated Barabási-Albert networks considered in figure 4.13: $N = 10^4$ (grey), $N = 10^5$ (magenta), $N = 10^6$ (blue).

and 3 do not have this disadvantage). Additionally, the algorithm for generating this type of networks (chapter 2) allows us to construct networks with a fixed order, but not with a given size. Thus, each realization usually has a different number of edges (which, of course, fluctuates around a mean value). So, the typical deviation of $\langle k \rangle$ is relatively large (except for $N \to \infty$), and consequently, the percolation results exhibit big error bars. Although we also explored the percolation properties of model 4, we refrain from presenting our findings, since no accurate conclusions could be extracted.

Degree-degree correlated scale-free networks. Assortative and dissortative mixing have a big impact on the percolation behavior of networks. We will see that the removal of a relatively small fraction of vertices (or edges) severely damage the structure of assortatively mixed networks. On the other hand, dissortative networks are distinctly more robust than uncorrelated networks. Recent studies related to the synchronousness and dynamical stability of networks have pointed out similar results: dissortative networks are more resistant to the effect of dynamical fluctuations than assortative networks [di Bernardo et al., 2005] and, assortative mixing by degree tends to destabilize networks [Brede and Sinha, 2005]. Interestingly, the study of the percolation thresholds shows the opposite tendency: dissortatively mixed networks undergo the phase transition before uncorrelated networks, however, strongly assortative networks seem to experience no transition.

Figure 4.15 shows the behavior of $\mathcal{M}$ for different degree-correlated Barabási-Albert networks; panel (a) corresponds to site percolation and panel (b) to bond percolation. The results are from networks with $N = 10^5$ and $L = 2 \cdot 10^5$ $(m = 2)$, and are averaged again over ten independent realizations of the corresponding algorithms. The two upper curves of both panels show the behavior of two of the many dissortative DS networks we investigated (see Xulvi-Brunet and Sokolov [2005] for more results): $p = 0.78$ (yellow), and $p = 1$ (dark green), where we have used the parameter p (the control parameter of the dissortative algorithm) for measuring the degree of dissortativity of the networks. The simulations demonstrate that, as the dissortativity grows, the robustness of the networks increases. This tendency, however, is only valid provided that the fraction of removed vertices (edges) is not too large; for concentrations of removed elements larger than $q \simeq 0.70 \simeq \bar{q}$ the networks tend to break up very rapidly. The uncorrelated network (dark blue curves of the pictures) evidences a robustness smaller than dissortative networks as long as $q \leq 0.70 \geq \bar{q}$, but, on the other hand, it disintegrates at larger concentrations. (The study of the percolation thresholds is shown in figures 4.16 - 4.21.) The other curves display the percolation behavior of the assortative correlated AS networks; from top to bottom, and using

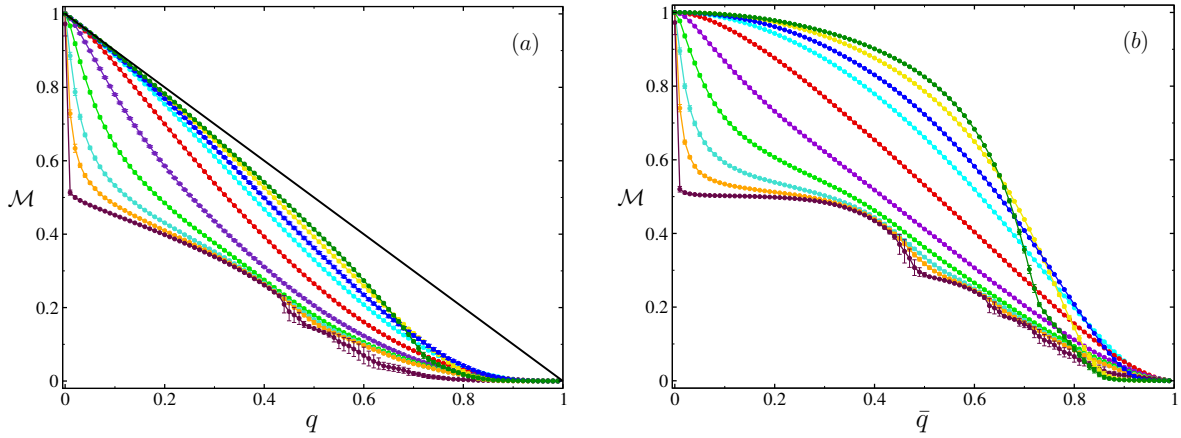

Figure 4.15: (a): Site percolation. Fraction of vertices $\mathcal{M}$ in the largest component depending on the fraction of vertices removed from the network q. (b): Bond percolation. Fraction of vertices $\mathcal{M}$ in the largest component depending on the fraction of edges removed from the network $\bar{q}$. The graphs compare the percolation behavior of Barabási-Albert networks with different mixing. From top to bottom: the dissortative networks $p = 1$ (dark green) and $p = 0.78$ (yellow), the uncorrelated version (blue), and the assortative networks $\mathcal{A} = 0.069$ (cyan), $\mathcal{A} = 0.221$ (red), $\mathcal{A} = 0.443$ (indigo), $\mathcal{A} = 0.640$ (green), $\mathcal{A} = 0.777$ (turquoise), $\mathcal{A} = 0.856$ (orange), and $\mathcal{A} = 0.913$ (maroon) (close to the maximal assortativity).

Eq. 3.12 to measure the assortativity: $\mathcal{A} = 0.069$ (cyan), $\mathcal{A} = 0.221$ (red), $\mathcal{A} = 0.443$ (indigo), $\mathcal{A} = 0.640$ (green), $\mathcal{A} = 0.777$ (turquoise), $\mathcal{A} = 0.856$ (orange), and $\mathcal{A} = 0.913$ (maroon) -close to the maximal assortativity. We note that the behavior of $\mathcal{M}$ gradually changes as the assortativity grows from the uncorrelated case (dark blue curve) to quite a different behavior when $\mathcal{A} \to 1$. The results, which obviously follow from topological differences between strongly assortative and uncorrelated networks, indicate a considerable decrease in the robustness of networks as assortativity increases.

Let us focus on the region where percolation transition (supposedly) takes place. Our aim will be to establish whether degree-degree correlations significantly change the critical threshold at which the phase transition happens or not. For this purpose, we consider models 1, 2, and 3 to which the AS, DS, and RS algorithms are applied. For each model, three different correlated versions are studied: a highly dissortative DS version ($p = 0.9999$), the uncorrelated version, and another strongly assortative version ($\mathcal{A} \simeq 0.777$). We examine networks of 10^4, 10^5, and 10^6 vertices in each considered model. Depending on the network's order (from the smallest to the largest), the simulation results are averaged over 20, 10, and 5 realizations, respectively. Our findings are plotted in figures 4.16 - 4.21. Each figure contains the following graphs: Graphs a, b, and c represent the behavior of $\mathcal{M}$, respectively, for the dissortative, uncorrelated, and assortative versions of each model. The insets display the same results on normal scales. For each model, the colors of the curves correspond to the following network characteristics: $N = 10^4$, dissortative (maroon); $N = 10^5$, dissortative (turquoise); $N = 10^6$, dissortative (orange); $N = 10^4$, uncorrelated (grey); $N = 10^5$, uncorrelated (violet); $N = 10^6$, uncorrelated (blue); $N = 10^4$, assortative (indigo); $N = 10^5$, assortative (yellow); $N = 10^6$, assortative (green); All six graphs (including the insets) compare the percolation behavior of the networks with order $N = 10^4$, $N = 10^5$, and $N = 10^6$. Graph d gives the percolation thresholds of the dissortative, uncorrelated, and assortative networks, together with the corresponding

power law fits; the fits and the obtained thresholds correspond to the networks of 10^6 vertices[8]. In graph e we compare the percolation behavior of the three different correlated versions (also for the largest networks, $N = 10^6$). Finally, panel f shows how the quantities $\mathcal{D}$ and $\mathcal{S}$ change as the removed fraction of elements -vertices or edges- is increased.

The main conclusion that we can immediately draw is that the percolation threshold tends to the larger values of q_c ($\bar{q}_c$) as the network changes from dissortative to assortative. The model in which this is most evident seems to be model 1. The pictures, both for site and bond percolation, certainly indicate a phase transition at $q_c^d \simeq 0.733 \simeq \bar{q}_c^d$ for the dissortative network and $q_c^r \simeq 0.834 \simeq \bar{q}_c^r$ for uncorrelated networks. It is remarkable that the site and bond percolation thresholds are approximately the same (the simulations show only a very small difference between them, namely $\bar{q}_c > q_c$). This characteristic also occurs in random graphs (although not in lattices). Note further that the threshold $q_c^r = 0.834$ coincides with the theoretical prediction of Eq. 4.15, which shows that the disagreement between this equation and our preceding results from the diverging scale-free networks is only valid when studying networks whose maximum degree rapidly grows with N [Molloy and Reed, 1995, 1998]. We also remark the fact that the largest cluster again becomes a tree above the percolation thresholds, as well as that the quantity $\mathcal{S}$ exhibits definite peaks at the fraction of q ($\bar{q}$) where phase transition occurs. The situation is different for the assortative networks; unlike the two previous cases in which the simulations indicate that the larger networks ($N = 10^5$ and $N = 10^6$) are already very close to the thermodynamic limit, figures (c) and (f) show a certain displacement of the curves as the network's order increases, particularly in the site percolation case. Note also that the behavior of $\mathcal{M}$ (panel c) exhibits no inflexion point, in contrast to the dissortative and uncorrelated cases. We conclude that, if the bond and site percolation transitions occur, then they must take place at values extremely close to $\bar{q}_c = 1 = q_c$. Finally, it is interesting to discuss the behavior of parameter $\mathcal{S}$ in this assortative case: $\mathcal{S}$ does not peak at large q and $\bar{q}$ but at very small values. The peaks at small q and $\bar{q}$ arise because, in the assortative case, the system breaks into a large component (containing a small but finite fraction of the network elements) and many other components composed of moderate, but not necessarily small degree vertices, when the fraction of removed elements is very small. These conclusions are confirmed by the study of the order distribution of components). The lack of peaks at large fractions of q and $\bar{q}$ may be interpreted as additional evidence for the non-existence of the corresponding percolation transitions. We note also that the largest cluster of assortative networks always contains cycles, even for very large fractions of the removed elements.

The findings of model 2 and 3 are qualitatively similar to those of model 1. The simulations indicate the existence of percolation thresholds for the uncorrelated and dissortative networks, both for site and bond percolation. However, the corresponding assortative versions do not seem to undergo any transition. As before, the threshold $q_c^r \simeq 0.715$ that we obtain for the uncorrelated model 2 lies close to the theoretical threshold (Eq. 4.15) $q_c \sim 0.69$. The difference could be due to the difficulty of experimentally estimating q_c: first, as a result of the large error bars of the simulations, second, because the system still seems far away from the thermodynamic limit, even for $N = 10^6$ (see, for example, panel f), and third, because the average first moment $\langle \bar{k} \rangle$ of the degree distribution clearly increases with N (in this case, $\langle \bar{k} \rangle (N = 10^4) = 1.344 \pm 0.012$, $\langle \bar{k} \rangle (N = 10^5) = 1.3465 \pm 0.0025$, and $\langle \bar{k} \rangle (N = 10^6) = 1.3478 \pm 0.0010$). With regard to model 3, it is interesting to point to the double peak which quantity $\mathcal{S}$ shows in the dissortative case. We will discuss this feature later. We remark again that our simulations indicate the existence of true percolation transitions (both for the site and bond percolation processes) in the case of

[8] We assume that the percolation thresholds of the infinite versions lie very close to those of the networks of order $N = 10^6$.

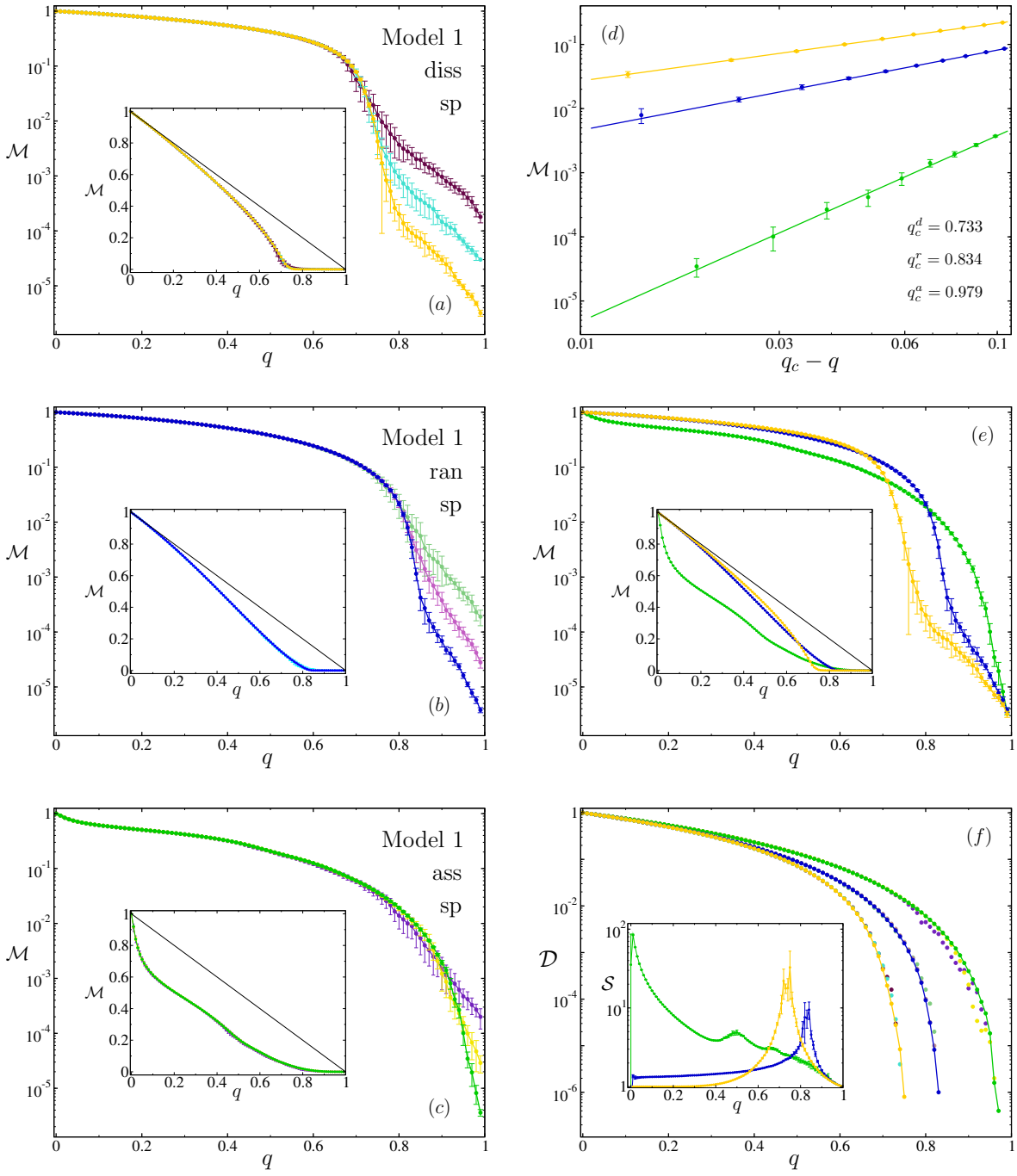

Figure 4.16: Model 1. Site percolation. (a), (b), and (c): $\mathcal{M}$ depending on q for the dissortative, uncorrelated, and assortative networks considered, respectively (see text). In each picture, the curves correspond, from top to bottom, to the networks of 10^4, 10^5, and 10^6 vertices. (d): $\mathcal{M}$, as a function of $q_c - q$ when $q \to q_c^-$, for the networks of 10^6 vertices, and the three mixed networks considered. Critical fractions: $q_c^d = 0.733$, dissortative (orange); $q_c^r = 0.834$, uncorrelated (blue); $q_c^a \simeq 0.979$, assortative (green). (e) Behavior of $\mathcal{M}$ versus q. Comparison between the assortative (green), uncorrelated (blue), and dissortative (orange) networks of order $N = 10^6$. (f): $\mathcal{D}$ (picture) and $\mathcal{S}$ (inset) as a function of q for the three mixed networks considered ($N = 10^6$).

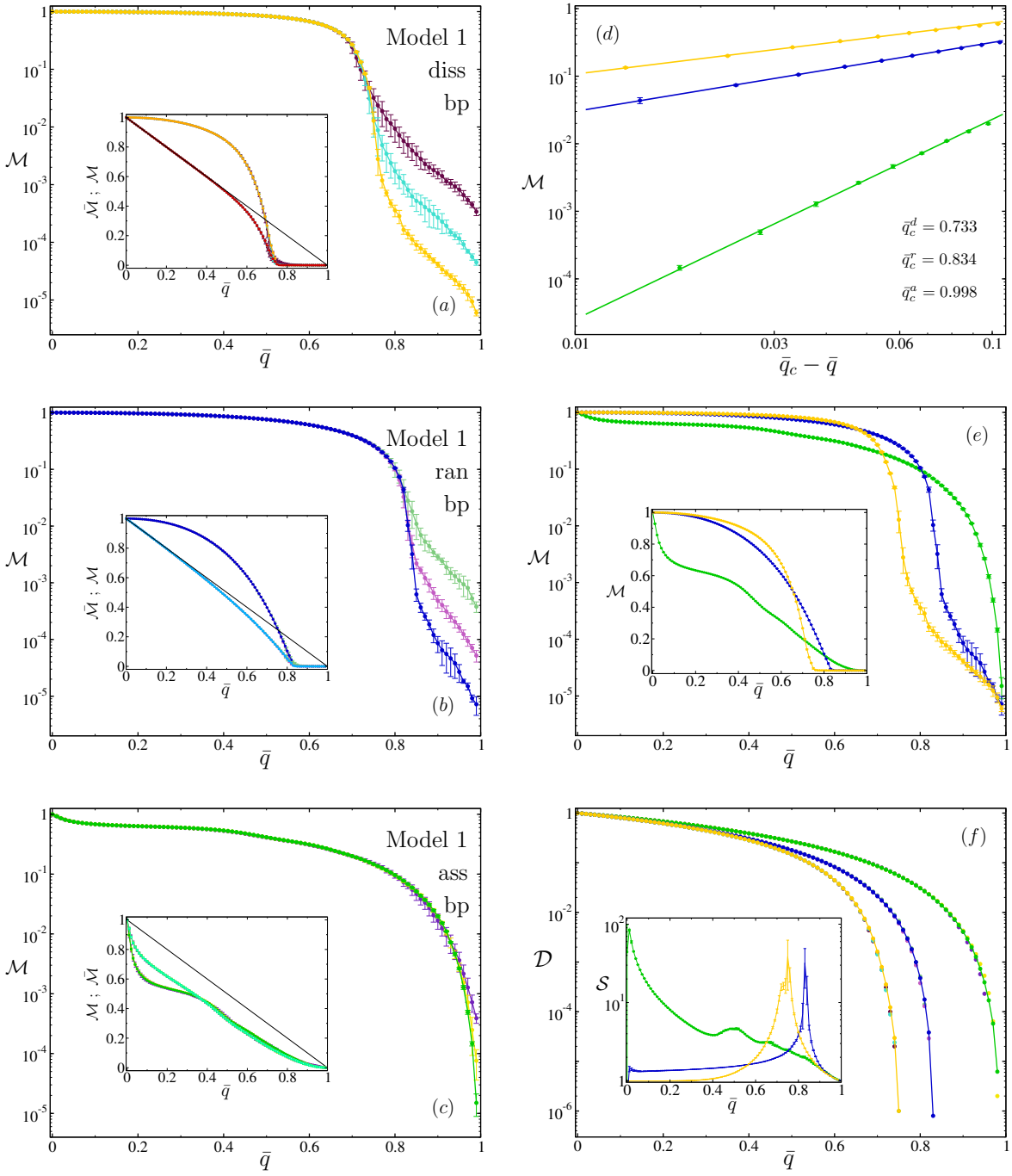

Figure 4.17: Model 1. Bond percolation. (a), (b), and (c): $\mathcal{M}$ depending on $\bar{q}$ for the dissortative, uncorrelated, and assortative networks considered, respectively (see text). In each picture, the curves correspond, from top to bottom, to the networks of 10^4, 10^5, and 10^6 vertices. (d): $\mathcal{M}$, as a function of $\bar{q}_c - \bar{q}$ when $\bar{q} \to \bar{q}_c^-$, for the networks of 10^6 vertices, and the three mixed networks considered. Critical fractions: $\bar{q}_c^d = 0.733$, dissortative (orange); $\bar{q}_c^r = 0.834$, uncorrelated (blue); $\bar{q}_c^a \simeq 0.998$, assortative (green). (e) Behavior of $\mathcal{M}$ versus $\bar{q}$. Comparison between the assortative (green), uncorrelated (blue), and dissortative (orange) networks of order $N = 10^6$. (f): $\mathcal{D}$ (picture) and $\mathcal{S}$ (inset) as a function of $\bar{q}$ for the three mixed networks considered ($N = 10^6$).

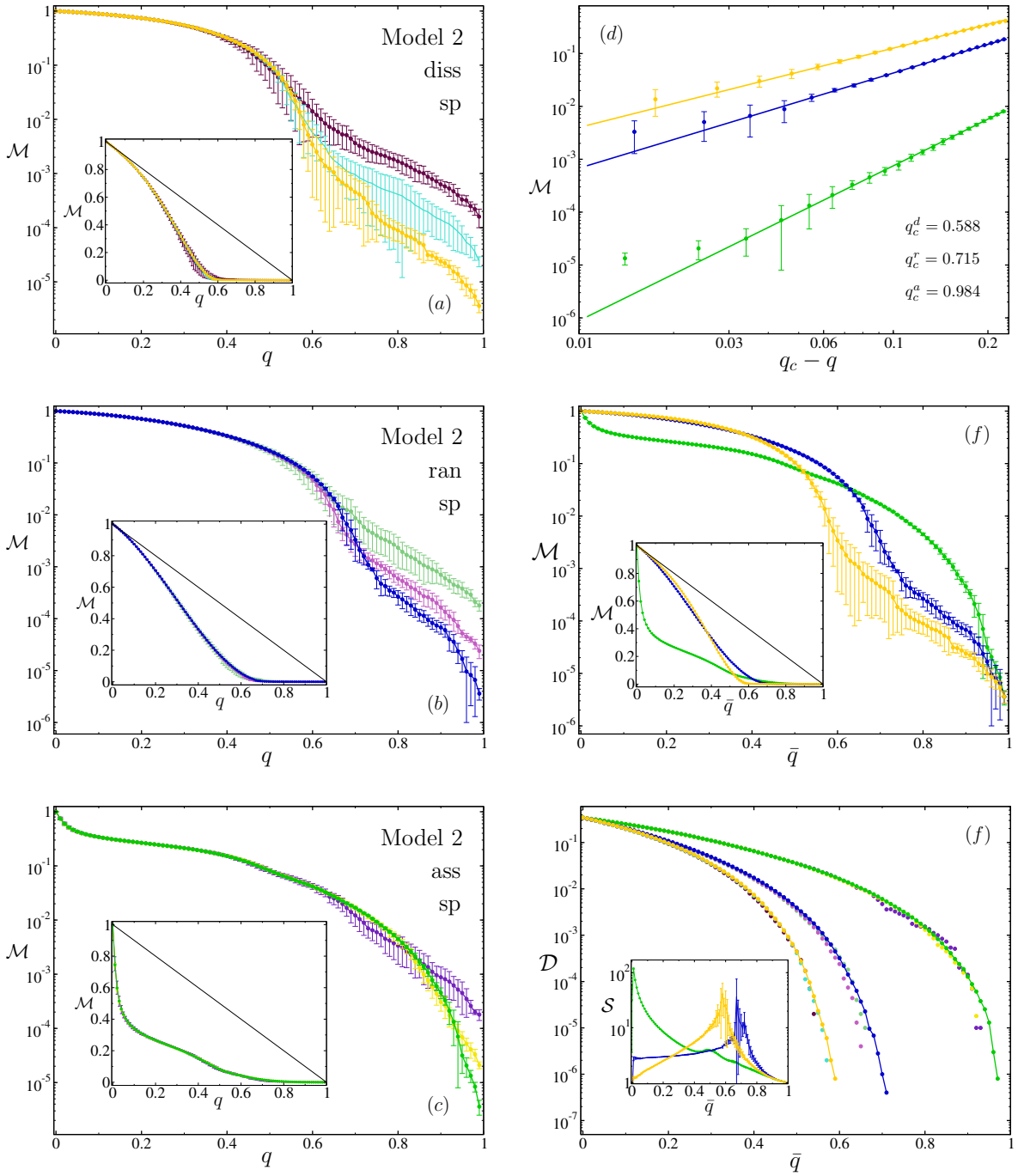

Figure 4.18: Model 2. Site percolation. (a), (b), and (c): $\mathcal{M}$ depending on q for the dissortative, uncorrelated, and assortative networks considered, respectively (see text). In each picture, the curves correspond, from top to bottom, to the networks of 10^4, 10^5, and 10^6 vertices. (d): $\mathcal{M}$, as a function of $q_c - q$ when $q \to q_c^-$, for the networks of 10^6 vertices, and the three mixed networks considered. Critical fractions: $q_c^d = 0.588$, dissortative (orange); $q_c^r = 0.715$, uncorrelated (blue); $q_c^a \simeq 0.984$, assortative (green). (e) Behavior of $\mathcal{M}$ versus q. Comparison between the assortative (green), uncorrelated (blue), and dissortative (orange) networks of order $N = 10^6$. (f): $\mathcal{D}$ (picture) and $\mathcal{S}$ (inset) as a function of q for the three mixed networks considered ($N = 10^6$).

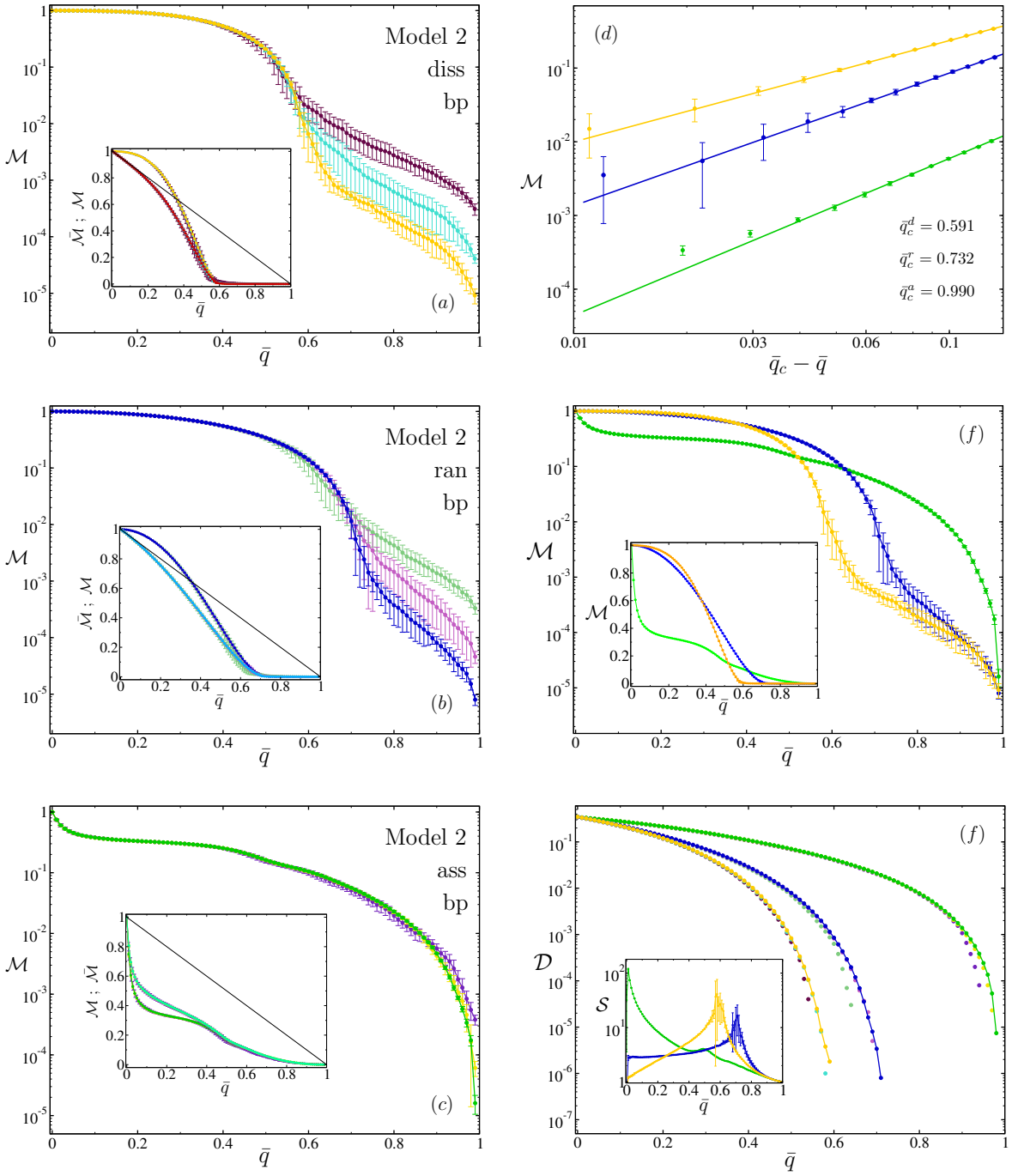

Figure 4.19: Model 2. Bond percolation. (a), (b), and (c): $\mathcal{M}$ depending on $\bar{q}$ for the dissortative, uncorrelated, and assortative networks considered, respectively (see text). In each picture, the curves correspond, from top to bottom, to the networks of 10^4, 10^5, and 10^6 vertices. (d): $\mathcal{M}$, as a function of $\bar{q}_c - \bar{q}$ when $\bar{q} \to \bar{q}_c^-$, for the networks of 10^6 vertices, and the three mixed networks considered. Critical fractions: $\bar{q}_c^d = 0.591$, dissortative (orange); $\bar{q}_c^r = 0.732$, uncorrelated (blue); $\bar{q}_c^a \simeq 0.990$, assortative (green). (e) Behavior of $\mathcal{M}$ versus $\bar{q}$. Comparison between the assortative (green), uncorrelated (blue), and dissortative (orange) networks of order $N = 10^6$. (f): $\mathcal{D}$ (picture) and $\mathcal{S}$ (inset) as a function of $\bar{q}$ for the three mixed networks considered ($N = 10^6$).

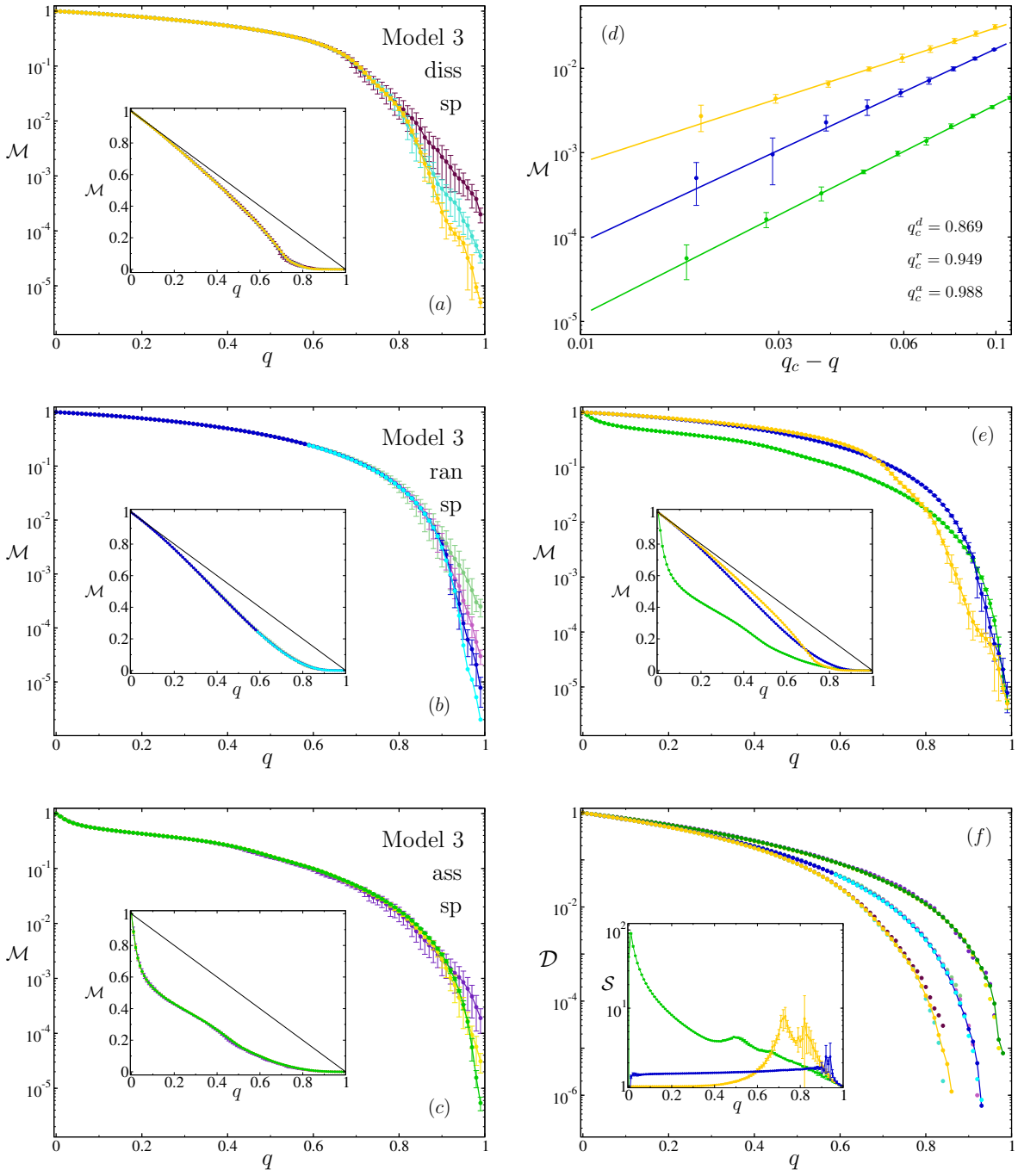

Figure 4.20: Model 3. Site percolation. (a), (b), and (c): $\mathcal{M}$ depending on q for the dissortative, uncorrelated, and assortative networks, respectively. In each picture, the curves correspond, from top to bottom, to the networks of 10^4, 10^5, and 10^6 vertices (except for cyan curves, which correspond to $N = 10^7$). (d): $\mathcal{M}$, as a function of $q_c - q$ when $q \to q_c^-$, for the networks of 10^6 vertices, and the three mixed networks considered. Critical fractions: $q_c^d = 0.868$, dissortative (orange); $q_c^r = 0.949$, uncorrelated (blue); $q_c^a \simeq 0.988$, assortative (green). (e) Behavior of $\mathcal{M}$ versus q. Comparison between the assortative (green), uncorrelated (blue), and dissortative (orange) networks of order $N = 10^6$. (f): $\mathcal{D}$ (picture) and $\mathcal{S}$ (inset) as a function of q for the three mixed networks considered ($N = 10^6$).

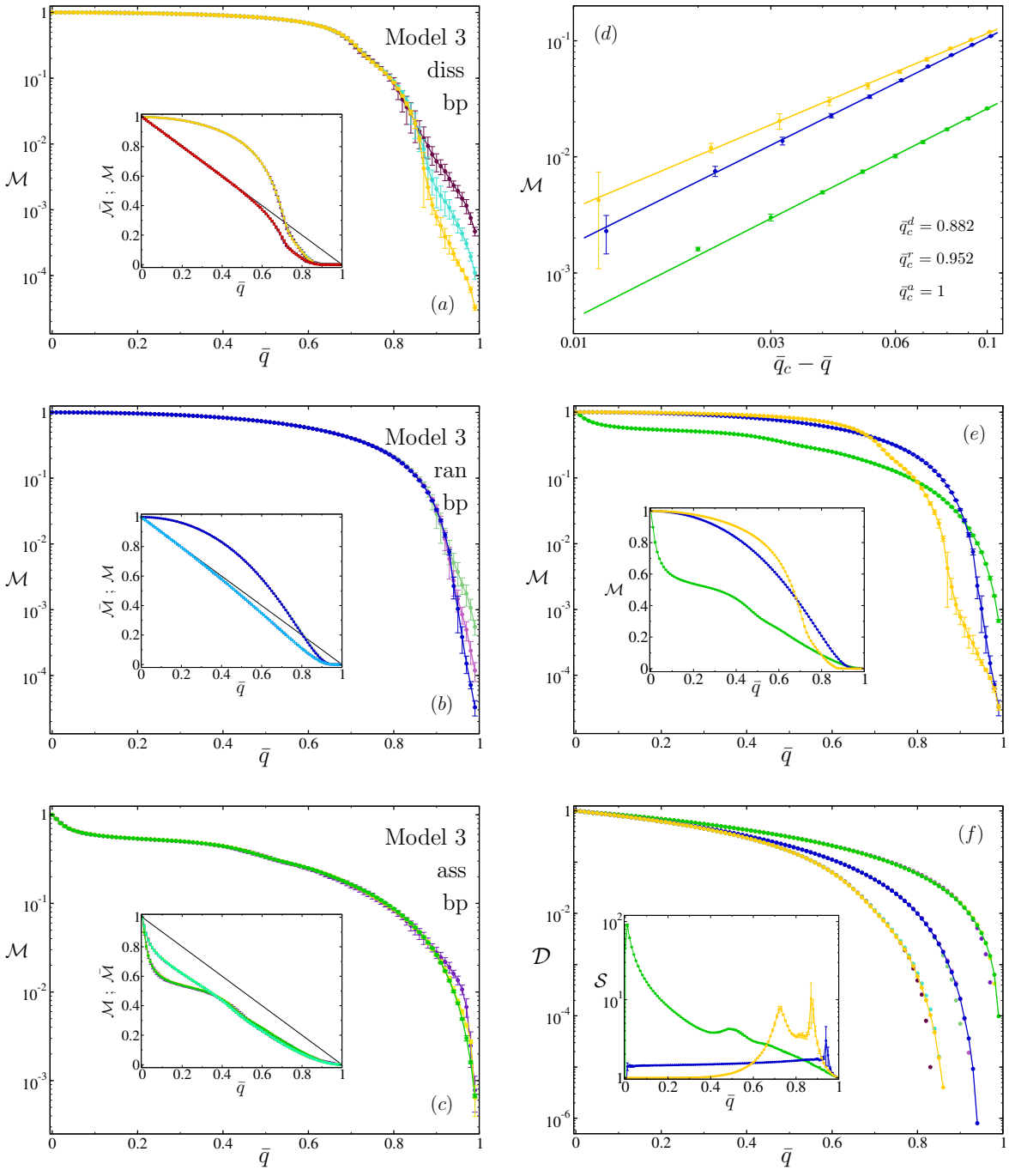

Figure 4.21: Model 1. Bond percolation. (a), (b), and (c): $\mathcal{M}$ depending on $\bar{q}$ for the dissortative, uncorrelated, and assortative networks considered, respectively (see text). In each picture, the curves correspond, from top to bottom, to the networks of 10^4, 10^5, and 10^6 vertices. (d): $\mathcal{M}$, as a function of $\bar{q}_c - \bar{q}$ when $\bar{q} \to \bar{q}_c^-$, for the networks of 10^6 vertices, and the three mixed networks considered. Critical fractions: $\bar{q}_c^d = 0.882$, dissortative (orange); $\bar{q}_c^r = 0.952$, uncorrelated (blue); $\bar{q}_c^a \simeq 1$, assortative (green). (e) Behavior of $\mathcal{M}$ versus $\bar{q}$. Comparison between the assortative (green), uncorrelated (blue), and dissortative (orange) networks of order $N = 10^6$. (f): $\mathcal{D}$ (picture) and $\mathcal{S}$ (inset) as a function of $\bar{q}$ for the three mixed networks considered ($N = 10^6$).

the uncorrelated network. Note that the system with $N = 10^6$ vertices already seems to be very close to the thermodynamic limit, which gives credibility to the statement. However, taking into account the inherent lack of precision of the simulations and the proximity of our results to unity, we refrain from concluding categorically the existence of a percolation transition. (For the particular case of site percolation, we carried out a very large simulation with a network of $N = 10^7$ vertices. It is represented by the cyan curves of figure 4.20. This curve shows a reasonably well distinguished point, $q \simeq 0.95$: $\mathcal{M}$ decreases as a power law below $q \simeq 0.95$, however, above this point the decrease is exponential. Assuming that the percolation behavior of this large network is similar to what we expect in the limit of $N \to \infty$ vertices, we could take this change of tendency as an indication for the existence of a phase transition.)

The big impact of degree correlations on the percolation behavior can be qualitatively explained as follows. Let us begin with the assortative correlations. We saw in chapter 3 that the principal characteristic of assortatively mixed networks is that they tend to form "communities" of vertices which have similar degree. This characteristic explains the fact that a small fraction of random breakdowns can gravely damage the global structure of assortative networks: Since different communities are poorly connected with each other (the degree of interconnection obviously depends on the degree of assortativity), the removal of a small fraction of vertices (or edges) is sufficient to break the originally connected network into separated "community-components". Most of them will be composed of low degree vertices, others will only contain moderately connected vertices, while one will concentrate on all vertices of large degree (which consequently means that it has a particularly large $\langle k \rangle$). Let us refer to this latter component as the "core" of the network. As the removal of elements increases, the clusters composed of small and moderate degree vertices tend to break up rapidly, while the "core" component, which is much more robust ($\langle k \rangle \gg 1$), perseveres as the only connected cluster (of course, gradually becoming smaller in size due to the loss of elements). The eternal persistence of this largest-core component results in an increment of q_c (or $\bar{q}_c$). Dissortative networks, on the other hand, have a more compact structure. Here, large degree vertices tend to be enclosed by low degree vertices so that hubs are not adjacent; low degree vertices operate as bridges which connect the large degree "island"-vertices. As a result of this, the system hardly contains small cycles; quite to the contrary, the system is filled with many large cycles (of course, this tendency depends on the degree of dissortativity). As we begin to eliminate vertices (or edges), the network tends to remain connected (because of the huge number of existing paths between any two vertices, which, in turn, is a direct consequence of the big number of large cycles) even for relatively large fractions of q (or $\bar{q}$). However, the largest component breaks up definitively when the large cycles disappear as a result of the removal of the "bridges" between hubs. Then, the network consists practically of a forest which quickly undergoes the percolation transition. Besides this, a careful analysis of the simulations shows that relatively large components begin to separate from the largest cluster at about q_c; for model 3, however, this begins by far earlier ($q \simeq 0.70$) compared to the point at which the largest cluster disintegrates completely ($q \simeq 0.875$). This peculiar behavior is especially interesting because, on the one hand, it demonstrates that phase transitions which take place in networks with diverging $\langle k^2 \rangle$ do not happen suddenly, and, on the other hand, it explains the appearance of the double peak found when analyzing the outcomes related to the quantity $\mathcal{S}$ for model 3. Quantity $\mathcal{S}$ increases as $q \to 0.70^-$ because intermediate components then begin to separate from the largest cluster. As q increases the order of these separated components decreases as the order of the largest component. This explains the existence of the first peak. The second peak is due to the definite breakdown of the largest component when the phase transition occurs -which coincides with the disappearance of cycles in the system.

Highly clustered scale-free networks. The percolation properties of highly triangulated

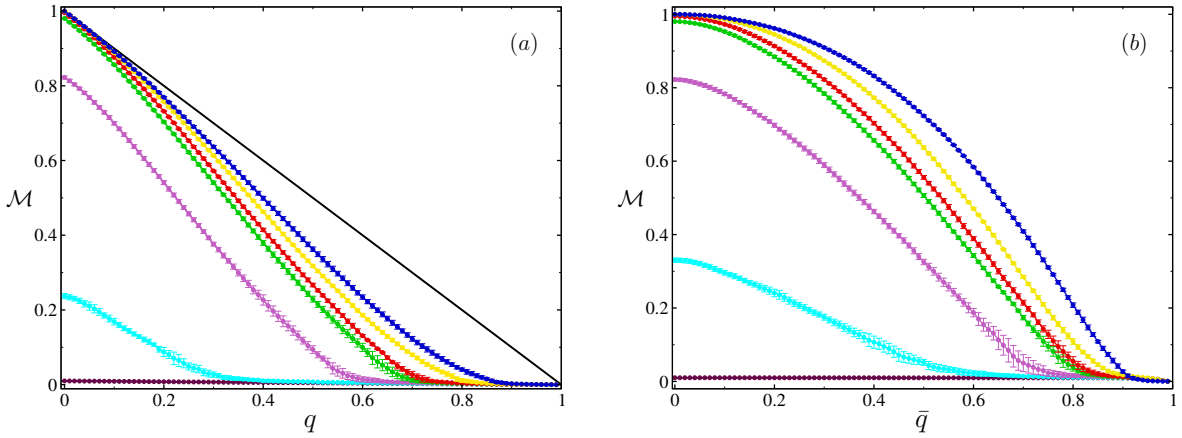

Figure 4.22: (a): Site percolation. Fraction of vertices $\mathcal{M}$ in the largest component depending on the fraction of vertices removed from the network q. (b): Bond percolation. Fraction of vertices $\mathcal{M}$ in the largest component depending on the fraction of edges removed from the network $\bar{q}$. The graphs compare the percolation behavior of highly triangulated Barabási-Albert networks. From top to bottom: $C = 0.0013$, corresponding to the uncorrelated version of the network (blue), $C = 0.0126$ (yellow), $C = 0.0695$ (red), $C = 0.1203$ (green), $C = 0.3833$ (violet), $C = 0.8463$ (cyan), and $C = 0.9958$, corresponding to the maximum reached value of C (maroon).

scale-free networks depend strongly on the degree of transitivity. Here, we explore the percolation behavior of different triangulated Barabási-Albert TS networks. We will see that the phase transition takes place at constantly diminishing fractions of removed elements (vertices or edges) as the mean clustering coefficient, C, grows. Thus, the situation seems to be similar to the one found for scale-free networks with underlying geography [Warren et al., 2002, Huang et al., 2005b]. These studies, combined with our results, could indicate that, independent of what mechanism increments the transitivity degree of the network, highly triangulated scale-free networks undergo a phase transition at finite concentrations.

Figure 4.22 shows the fraction of vertices $\mathcal{M}$ in the largest component versus the fraction q (panel a, site percolation) and $\bar{q}$ (panel b, bond percolation) of elements removed from the network. The graph compares the percolation behavior of several networks having different degree of transitivity. From top to bottom, the curves correspond to networks with $C = 0.0013$ (minimum value of the mean clustering coefficient, corresponding to $p = 0$, and therefore, to the uncorrelated version of the network), $C = 0.0126$, $C = 0.0695$, $C = 0.1203$, $C = 0.3833$, $C = 0.8463$, and $C = 0.9958$ (corresponding to the highest reached value of C). The order of the networks is $N = 3 \cdot 10^4$ and their size $L = 6 \cdot 10^4$. Especially interesting are the four top curves (that is, the ones verifying $C \leq 0.12$), which correspond to triangulated and initially fully connected scale-free networks, just like real-world networks. Note that the percolation threshold decreases as C grows.

The existence of the corresponding thresholds is explicitly pointed out by the results shown in figure 4.23. Here, we focus only on two values of the mean clustering coefficient: $C \simeq 0.0184$ (panel a) and $C \simeq 0.384$ (panel b). Additionally, we treat two orders, namely $N = 3 \cdot 10^4$ and $N = 3 \cdot 10^5$ ($L = 2 \cdot N$). The figure shows the behavior of $\mathcal{M}$ as a function of the removed fraction of vertices, q, (lower curves of both panels) and the removed fraction of edges, $\bar{q}$ (upper curves). The results exhibit the characteristic behavior of phase transitions: below a certain value of q

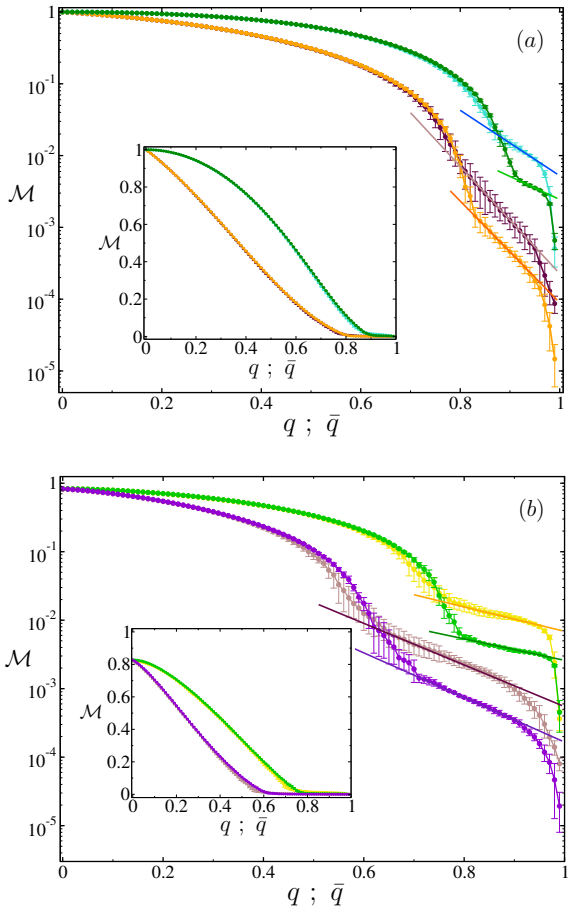

Figure 4.23: Fraction of vertices in the largest component, $\mathcal{M}$, as a function of the fraction q of removed elements (site percolation) or $\bar{q}$ (bond percolation). (a): All curves have $C \simeq 0.0184$. Turquoise, $N = 3 \cdot 10^4$, bond percolation. Dark green, $N = 3 \cdot 10^5$, bond percolation. Maroon, $N = 3 \cdot 10^4$, site percolation. Orange, $N = 3 \cdot 10^5$, site percolation. (b): All curves have $C \simeq 0.384$. Yellow, $N = 3 \cdot 10^4$, bond percolation. Green, $N = 3 \cdot 10^5$, bond percolation. Braun, $N = 3 \cdot 10^4$, site percolation. Indigo, $N = 3 \cdot 10^5$, site percolation. Note the logarithm scale on the y-axis in the pictures. The insets represent the same curves on normal scales.

(or $\bar{q}$ in bond percolation) all the different networks exhibit the same quantitative behavior[9], while, above this critical fraction, the results distinctly demonstrate the typical tendency of $\mathcal{M}$ to vanish. Notice further the exponential decrease of $\mathcal{M}$ above the critical concentration (in the figures, this exponential decrease is remarked with straight lines). Additional evidence found shows that the decrease of $\mathcal{M}$ below the critical concentration follows a power law in all cases. Figure 4.24 shows, on double logarithmic scales, the corresponding fits of the site percolation data. The fits are carried out on the largest networks considered, $N = 3 \cdot 10^5$. They give the critical thresholds $q_c = 0.820$ ($C \simeq 0.0184$) and $q_c = 0.640$ ($C \simeq 0.384$). The perfect agreement of these simulation results with the scaling assumption supports the existence of true percolation transitions in strongly triangulated scale-free networks.

The results shown in figures 4.22 - 4.24 follow from the peculiar structure of the triangulated TS networks, which is somehow a mixture between current networks and non-regular lattices. Sure enough, when $p = 0$, the TS algorithm creates an ordinary network without correlations

[9]In fact, the curves do not perfectly coincide, which could imply that the considered networks are still far away from the thermodynamic limit. The reason for the non-coincidence lies, however, in the fact that the mean clustering coefficients, C, are not exactly the same. The problem resides in the difficulty to generate networks with the same C but different order N. In contrast to what occurs with our assortative and dissortative algorithms, for which the same value of p generates topologically equivalent networks irrespectively of their order, when applying the TS algorithm on networks of different order the value of p must be changed in order to obtain the same C. The effect is due to the fact that C tends to be smaller in large networks than in small networks: the larger the network the smaller is the probability that triangles are formed. So, we tried many values of p until the generated versions had approximately the desired C.

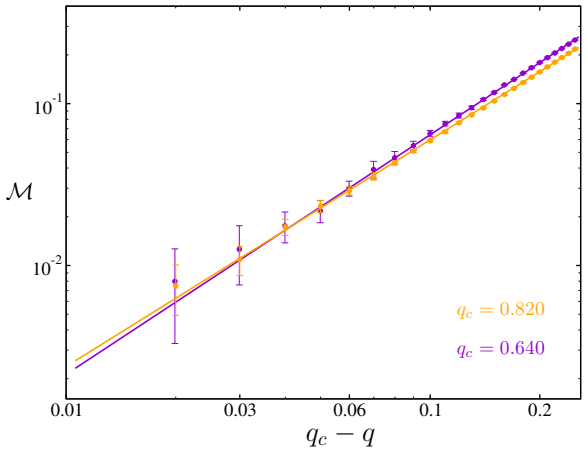

Figure 4.24: Fraction of vertices in the largest component, $\mathcal{M}$, as a function of $q_c - q$ when $q \to q_c^-$. The results correspond to networks of $3 \cdot 10^5$ vertices. $C \simeq 0.0184$, $q_c = 0.820$ (indigo); $C \simeq 0.384$, $q_c = 0.640$ (orange). Note the double logarithmic scale.

between vertices, and where all vertices are connected by means of relatively short paths. However, when p is large, the algorithm creates a network where vertices are grouped in small highly connected clusters, and the distance between vertices of different groups is usually large; as in lattices, vertices of the same cluster are typically neighbors, and vertices of different clusters are somewhat far away from each other. This feature is thus also reproduced by assortative networks. However, they seem to exhibit no percolation transitions. The crucial point to understand the different percolation behavior is the order of clusters: While "communities" in assortative networks usually contain many vertices, highly connected clusters are typically very small in triangulated networks ($N_{cl} \sim k$). For this type of networks no "core"-component containing a non-negligible fraction of vertices exists. It is true that there are many extremely connected clusters in the network, but, firstly, the fraction between the order of the clusters and N vanish when $N \to \infty$, and secondly, they are poorly interconnected. Therefore, when a large enough fraction of elements is eliminated, it breaks up totally, to the effect that no spanning component survives. This aspect of triangulated networks is reproduced in figure 4.22. Note also, that at concentrations much larger than the critical q_c at which the phase transition takes place figure 4.23 shows an abrupt decrease of $\mathcal{M}$ when $q \simeq 1$. This is due to the definitive breakdown of the larger well-connected clusters (nevertheless, containing a negligible fraction of vertices).

It is not surprising that triangulated networks undergo a true percolation transition. We already saw in section 3.3 that large TS networks can be described as uncorrelated pseudographs of supervertices. We also saw that, if the original network is scale-free $P(k) \sim k^{-\beta}$, then the supervertex degree distribution $\hat{P}(d)$ falls approximately as a power-law $\hat{P}(d^{-\gamma})$, and the corresponding exponent γ decreases as the transitivity grows. Moreover, the exponent γ takes the value $\gamma \simeq \beta - 1$ when $p \simeq 1 - 1/3L$, which entails $\gamma > 3$ for the most interesting networks. Additionally, if the network is large, this all still occurs even if the mean clustering coefficient of the network is small. When we apply a bond or site percolation process to a highly triangulated network, the supervertices will presumably not break during the process since they appear to be very robust. However, the structure of external edges connecting the supervertices between themselves will without break because $\hat{P}(d)$ decays much more rapidly than a power-law of exponent $\gamma = 3$. (Bear in mind that uncorrelated networks for which the second moment of the degree distribution is finite undergo a true percolation transition, according to Eq. 4.15). The effect that the percolation transition takes place at unceasingly smaller concentrations as the transtivity grows can also be explained: As C grows the exponent γ governing the degree distribution of supervertices increases progressively, which, according to Eq. 4.15, causes the

transition to occur repeatedly at smaller fractions of removed elements. Triangulated networks are therefore networks that can experience a true percolation transition if the mean clustering coefficient is large enough, even if they have a degree distribution with a diverging second moment.

Chapter 5

Evolving networks under geographical constraints

In many networks such as transportation and communication networks, the physical distance between vertices is certainly a relevant parameter. Many communication network devices have a short radio range, contagious diseases do not spread uniformly across territories [Sander et al., 2002, 2003, Hufnagel et al., 2004], and most people have their friends in their geographical neighborhood. The density of people in cities decreases exponentially from the center [Makse et al., 1998, Andersson et al., 2002], and the interstate road network has only very short edges, in the order of 10km to 100km [Gastner and Newman, 2004]. In addition, examples of real-world networks suggest that when long-range edges exist, they usually connect hubs. Important examples of this are the Internet, in which connections are made by means of physical cables with different lengths, or airline networks, where long connections point preferably to big airports [Yook et al., 2002, Gastner and Newman, 2004]. For all these networks the connection probabilities are presumably both a function of what is technologically desirable and what is geographically feasible. Thus, even if some networks are defined without reference to an embedding space, it is not the case for most real-world networks.

Networks in which vertices are placed in a physical space are usually called "spatial" networks, and their principal characteristic is that the connection probabilities among their vertices depend on the Euclidean distance. In most cases, such as social interactions or transportation networks, the range of interaction is limited, which is explained by the fact that there is a cost associated with long-range connections. If the cost of a long-range edge is high, most of the connections starting from a given vertex link to the nearest neighbors; on the other hand if this cost is low the edges can connect to distant vertices. Several models have been proposed in the past few years in order to reproduce the statistical properties of spatial networks. Most of them combine the preferential attachment, which is widely accepted as the probable explanation for the power law distribution seen in many networks, and distance effects [Xulvi-Brunet and Sokolov, 2002, Manna and Sen, 2002, Barthélemy, 2003, Sen and Manna, 2003, Manna and Kabakçioğlu, 2003, Jost and Joy, 2002]. Typically, this leads to some crossover away from scale-free behavior when the distance constraint is sufficiently strong. A different approach to understand the interplay of geography and topology has been to consider ways in which a scale-free network can be embedded in Euclidean space and where the long-range connections of a vertex of degree k_i are restricted within the limits of a radius function $r(k_i)$ containing the dependence with the distance [Rozenfeld et al., 2002, Warren et al., 2002, Herrmann et al., 2003, Dall and Christensen, 2002, Yang et al., 2004, ben Avraham et al., 2003]. Other spatial networks which have recently been studied are the so-called Apollonian networks [Doye and Massen, 2005a, J. S. Andrade et al.,

2005, Massen and Doye, 2005], which use space-filling disk packings to provide a useful model for energy landscape networks [Doye, 2002, Doye and Massen, 2005b]. Very recently Barrat et al. [2005] and Masuda et al. [2005] have discussed the interrelation between weight dynamics and spatial constraints, showing that it is a key ingredient for understanding the formation of real-world weighted networks.

In this chapter we present two models which take this metrical aspect into account. Our goal is to demonstrate that geography plays a fundamental role in the emerging properties of evolving networks. The first model is a one-dimensional extension of the Barabási-Albert model which penalizes long-range connections on behalf of the connections jointing neighboring vertices. The second model combines in a selective way the tendency of vertices to connect by means of short-range connections (to minimize the cost) and the tendency to connect to the more "attractive" vertices of the network. We show that geographical restrictions induce spontaneously the appearance of large values of clustering (which overcomes the principal problem of most models of scale-free networks), and reproduces at the same time other real-world properties like the small-world effect and the scale-free character. In addition, the interdependence between geography and vertex attractiveness seems to be enough to explain the degree-degree correlations found in many (technological) networks.

5.1 Preferential attachment with disadvantaged long-range connections

In this section we introduce a simple spatial model which is based, like the genuine Barabási-Albert construction, on growth and preferential attachment. Here, however, the probability of connecting two vertices depends not only on the number of connections that the vertices already have, but also on the distance existing between them. That is, we treat an emerging network in a metrical space. To be more precise, in this emerging network the probability that a newly introduced vertex n connects to a previously existing vertex i is proportional to the number k_i of the already existing connections of vertex i (preferential attachment prescription), and where long edges are disadvantaged, because we impose that this probability also depends on the Euclidean distance d_{in} between vertices n and i as $d_{in}^{-\alpha}$, (clearly, a "scale-free" function), with $\alpha > 0$. Based on extensive numerical simulations of a one-dimensional situation we show that even if the length penalties are mild, the model exhibits properties that differ strongly from those of the usual scale-free networks. Thus, the corresponding degree distribution function $P(k)$ depends strongly on α. In particular, we show that for $\alpha < 1$ the behavior of $P(k)$ is similar to the behavior of the Barabási-Albert model without penalties, so that asymptotically $P(k) \propto k^{-3}$, while for $\alpha > 1$ the behavior of $P(k)$ is well described by a stretched exponential $P(k) \propto exp(-bk^{\gamma})$, with the power γ depending on α. The overall structure of the emerging network preserves its small-world nature even at large -and probably at all- α-values, and also presents large clustering coefficients.

We start from a one-dimensional lattice of L sites, spaced by a unit distance, and apply cyclic boundary conditions. We will let our network grow on this structure, so that each lattice site is a possible location of a network's vertex. We denote by n_i the position in the lattice of node i. The distance d_{ij} between any two nodes i and j is defined as

$$d_{ij} = \min\{|n_i - n_j|, (L - |n_i - n_j|)\} \ . \tag{5.1}$$

Let us now construct the network. First, we choose randomly an even number m_0 of sites from the lattice and we bind them in pairs with one edge each. This will be our initial condition.

That is, at $t = 0$, our network will consist of m_0 vertices connected in pairs. As in the Barabási-Albert model, we will add at every time step a new vertex to our network (linear growth). We proceed according to the following rule: at every time step, we choose at random a free site of our lattice, and there pose the new vertex. This new vertex is then connected by means of m edges ($m \leq m_0$) with m different vertices already present in the network. After t time steps the algorithm results in a network with $t + m_0$ vertices and $mt + m_0/2$ edges. In contrast with the Barabási-Albert model, the probability Π for the new vertex n to be connected to an old one i will depend not only on the number of edges k_i, which i already possesses, but also on the distance d_{in} between them,

$$\Pi(k_i, d_{in}, \alpha) = \frac{k_i \cdot d_{in}^{-\alpha}}{\sum_j k_j \cdot d_{jn}^{-\alpha}} \ . \tag{5.2}$$

Here the sum in the denominator extends over all vertices in the system except the newly introduced one, and α is a real non-negative parameter describing the distance penalties. For large α, the probability of a connection between two distant vertices is very small; on the other hand, for a very small α the probability is almost independent of the distance. In the case $\alpha = 0$, our model reduces to the genuine scale-free one. Note that our model is also to some extent scale-free: the connection probabilities depend only on the relative distances. (The initial condition is slightly different from that of Barabási and Albert, where the initial m_0 vertices are not connected: in our case all vertices introduced at $t = 0$ have exactly one edge, which allows us to use Eq. 5.2 from the very beginning. This simplifies the algorithm, since we do not have to distinguish between the initial and further steps. The only difference with the genuine Barabási-Albert construction is that, instead of mt edges, at time t $mt+m_0/2$ edges are present. The asymptotic behavior of both models for $t \to \infty$ is, however, the same.)

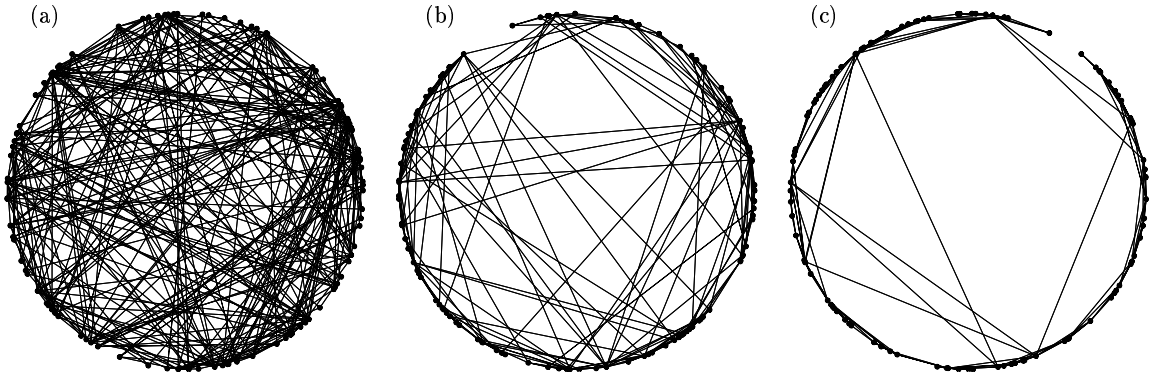

Figure 5.1: Small networks generated using the algorithm prescription, Eq. 5.2, with different values of α: (a) $\alpha = 0$, (b) $\alpha = 1.5$, and (c) $\alpha = 15$. All three examples have 300 edges, $L = 10^6$, $N = 105$, and $m = 3$. Note the change in the appearance of the networks. The network (a) is a genuine Barabási-Albert scale-free construction while (c) strongly resembles the Watts and Strogatz small-world network.

Three examples of evolving networks of this kind are given in figure 5.1. Here $m = 3$, $L = 10^6$, $N = 105$, and $m_0 = 6$ (so that all three networks have exactly 300 edges). Three different values of α were used: $\alpha = 0$ (Barabási-Albert model), $\alpha = 1.5$ and $\alpha = 15$. Note that increasing the value of α leads to fundamental changes in the topology of the network. Figure

5.1 (a) corresponds to a genuine Barabási-Albert construction and exhibits a lot of long edges connecting distant sites. On the other hand, only few of such edges are present in figure 5.1 (c); for such large values of α the new vertex always connects to its immediate neighbors. In fact, all long connections are produced at the beginning of the network's construction (t small), when the neighbors are distant because of the low number of randomly placed vertices. The cost associated with edges is hidden in parameter α.

In our further simulations we use a lattice of $L = 2 \cdot 10^7$ sites; the maximum number of nodes is $N = 2 \cdot 10^5$. All simulation results are based on the average over ten realizations of this structure. The error bars in the next figures correspond to this ensemble average only. The simulations are done for several values of α and for two values of m, i. e., the number of incoming edges: $m = 1$ and $m = 3$; the initial condition is $m_0 = 2m$.

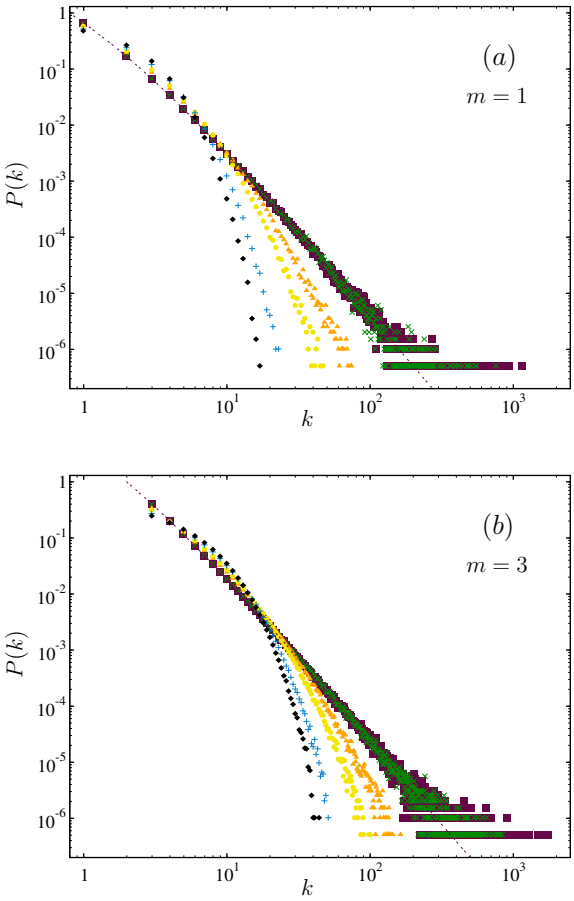

Figure 5.2: Degree distribution $P(k)$ arising from our spatial model. The pictures show the degree distribution for different values of m and α. The values of m are $m = 1$ (a), corresponding to a tree, and $m = 3$ (b). The values of α are $\alpha = 0$ (squares), $\alpha = 0.8$ (crosses), $\alpha = 1.5$ (triangles), $\alpha = 2$ (filled circles), $\alpha = 5$ (pluses), and $\alpha = 45$ (diamonds). The dashed lines correspond to the theoretical curve for the scale-free model, Eq. 1.6. All curves can be successfully fitted by the stretched-exponential function: $P(k) = a \exp(-bk^\gamma)$, where a, b and γ depend on α and m. For $\alpha < 1$ the degree distribution $P(k)$ follows the Eq. 1.6. The curves correspond to an emerging network of $N = 2 \cdot 10^5$.

The degree distribution of the scale-free Barabási-Albert model decays as a power law $P(k) \sim k^{-\gamma}$, with $\gamma = 3$. Let us now discuss how this distribution changes if the long-range connections are penalized. In figure 5.2 we plot the degree distribution $P(k)$ resulting from our model for different values of α on double logarithmic scales. The results show that for all $0 < \alpha < 1$, no important differences with the original Barabási-Albert construction ($\alpha = 0$) can be detected; Certain is that the asymptotic behavior of $P(k)$ is well described by $P(k) \sim k^{-\gamma}$. (The distributions seem to be almost identical. However, small, but statistically significant, deviations can be detected for small k values.) At $\alpha \simeq 1$ the degree distribution exhibits a pronounced change in its behavior and ceases being a power law. The analysis of the simulations suggests that the corresponding mathematical expression for the case $\alpha > 1$ could be a stretched-exponential

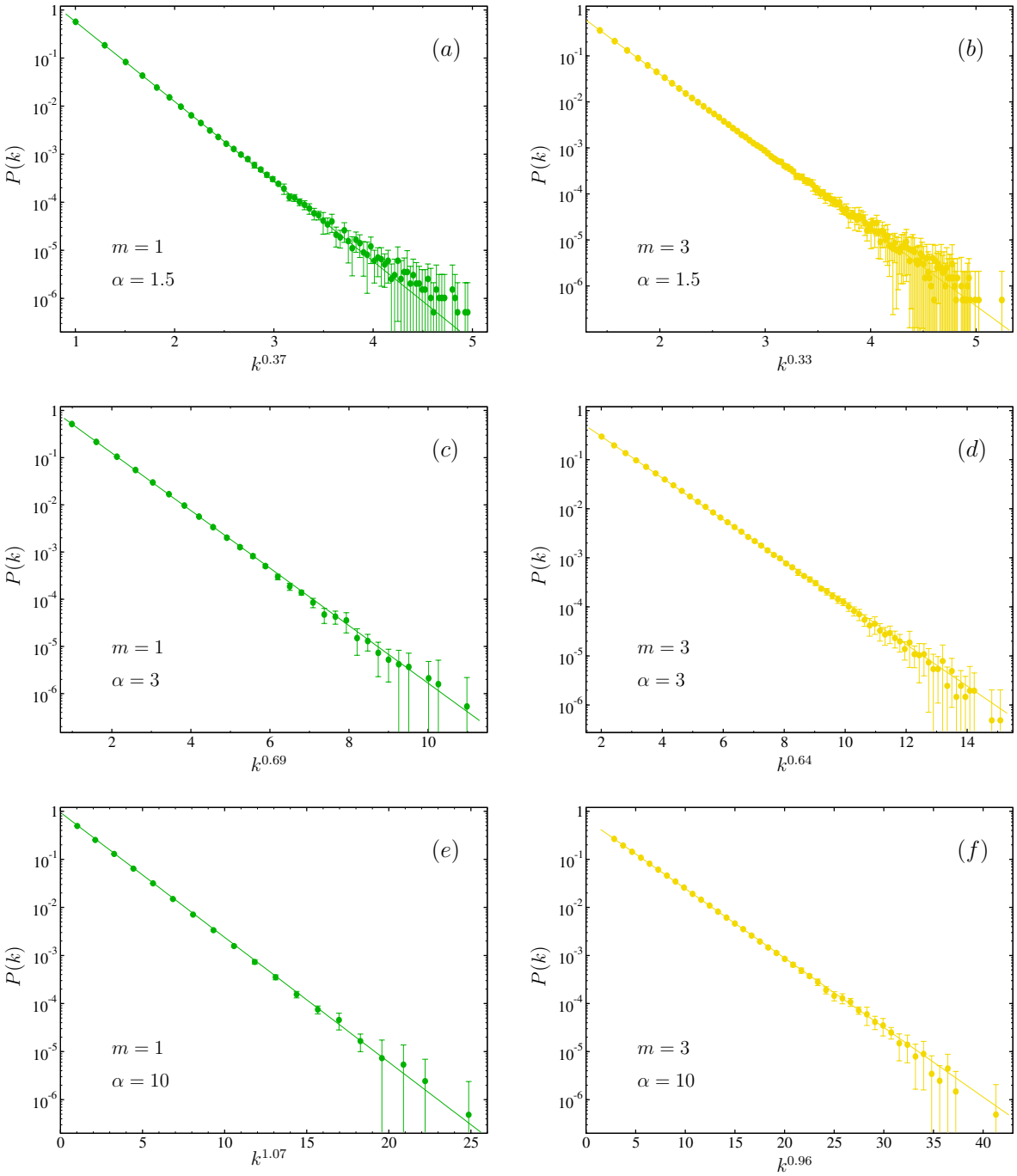

Figure 5.3: Shown is the degree distribution $P(k)$ of our spatial model based on preferential attachment with disadvantages long-range edges as a function of k^γ, where γ is the output of a fit of our data using Eq. 5.3, a stretched-exponential of the form $P(k) = a \exp(-bk^\gamma)$. The parameters are: (a) $m = 1$, $\alpha = 1.5$, and $\gamma = 0.37$; (b) $m = 3$, $\alpha = 1.5$, and $\gamma = 0.33$; (c) $m = 1$, $\alpha = 3$, and $\gamma = 0.69$; (d) $m = 3$, $\alpha = 1.5$, and $\gamma = 0.34$; (e) $m = 1$, $\alpha = 10$, and $\gamma = 1.07$; (f) $m = 3$, $\alpha = 10$, and $\gamma = 0.96$. The pictures assess the correctness of Eq. 5.3. Simulation results corresponding to other parameter values also show that such a fit is really excellent. We note that a fit of the data with a damped power law (see text for details) also offers good results for large α, but they are not as good as for small α, $1 < \alpha < 3$. All data correspond to evolving networks having $N = 2 \cdot 10^5$ vertices.

function of the form

$$P(k) = a \exp\left(-bk^{\gamma}\right) , \tag{5.3}$$

where the parameters a, b, and γ depend on α and m. To obtain the values of these parameters and to estimate the quality of this fitting function, we fit the data to Eq. 5.3 using the nonlinear least-squares Levenberg-Marquardt algorithm [Flannery et al., 1985], taking into consideration the error bars resulting from ten realizations of each situation. The data are re-plotted together with the outcomes of the fits in figure 5.3 on scales in which the fitting function, Eq. 5.3, is represented by a straight line. One takes k^{γ} as the abscissa and $\ln P(k)$ -note the logarithmic scale- as the ordinate of the graph. Figure 5.3 proves that such a fit is surprisingly good. The values of the exponent γ are shown as a function of α ($\alpha > 1$) in figure 5.4 for the two different situations corresponding to $m = 1$ and $m = 3$. We see that γ monotonically grows with α, and that the dependences for $m = 1$ and $m = 3$ differ, i. e., the $\gamma(\alpha)$ dependence is non-universal.

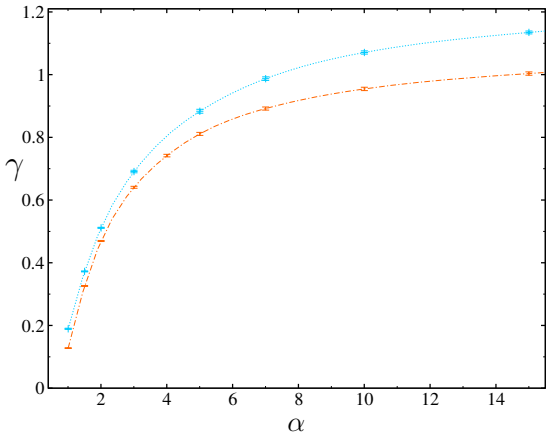

Figure 5.4: The parameter γ as a function of α. The upper dependence corresponds to $m = 1$, and the lower one to $m = 3$. The lines are drawn as a guide of eyes.

Newman [2001b] points out that many real world networks exhibit degree distributions which decrease as exponentially damped power laws $P(k) = ak^{\gamma} \exp(-bk)$, Eq. 1.3. This equation describes curves having a shape very similar to our stretched-exponential function, Eq. 5.3. We also tested this damped power law function, Eq. 1.3, as possible a fit of our simulation results and found out that it actually is a good fit for larger α values, but is definitely inferior to our fit, Eq. 5.3, for $1 < \alpha < 3$. Interestingly, Eq. 5.3 is related to expressions containing the gamma function which have been found in different theoretical approaches of evolving networks with non-linear preferential attachment [Krapivsky and Redner, 2001, Dorogovtsev and Mendes, 2001a]. Certain is that the similarity between Eq. 5.3, derived from our spatial approach, and Eq. 1.3, corresponding to real networks, could indicate a real influence of distance on the emerging structure of networks.

A growing network with disadvantaged long edges is quite an interesting construction. For large α, a strong correlation between the age of the connection and its length exists. Old connections, made when the nodes were sparse, are typically long, while younger connections get shorter and shorter, since more vertices in the immediate vicinity of a newly introduced vertex can be found. The simulations show that for large values of α, the vertices are almost surely connected to their nearest neighbors. On the other hand, the old, long-range connections are of great importance for the overall topology of the lattice, since they guarantee that for any α the network is a small world. In figure 5.5 we plot the average path length of the networks for the two different values $m = 1$ and $m = 3$ as a function of the network size N. Again, the error bars of the figure correspond to the mean path lengths over ten realizations of the network.

We see that the average path length of the networks grows linearly in $\ln N$, i. e., in accordance with small-world behavior. This behavior is preserved for all tested values of α; the largest value tested was $\alpha = 45$, which, for $m = 1$, ensures almost with certainty the connection of a newly introduced vertex to its geographical neighbors. The high-α networks closely resemble the simple small-world constructions [Watts and Strogatz, 1998]. Note also the similarity between this figure and figure 3.13 (b), corresponding to our assortative AS algorithm; it seems that "large" small world networks could frequently be realized in the real world.

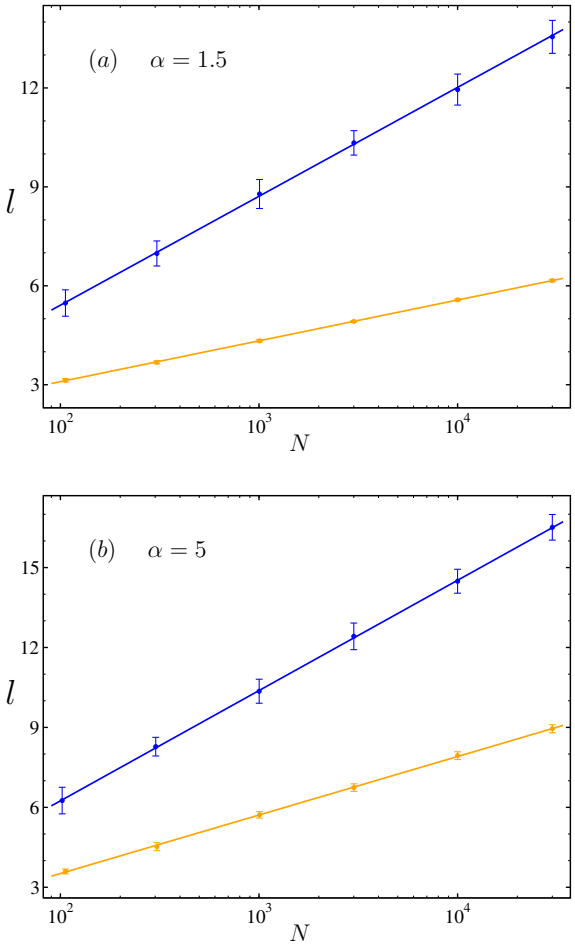

Figure 5.5: The shortest average path length l of a network as function of the number of vertices N. Panel (a) corresponds to $\alpha = 1.5$ and panel (b) corresponds to $\alpha = 5$. The upper lines in each panel are those for $m = 1$, the lower curves correspond to $m = 3$. Note the logarithm scale. We see that as m grows the average path length decreases.

The behavior of the clustering coefficient of the model also offers very interesting properties. Figure 5.6 shows how the clustering coefficients C and $C(k)$ change as the distance effects gain significance. In panel (a) we plot the mean clustering coefficient C as function of α, both on a logarithm scale (normal-size picture) and on a linear scale (inset) in the ordinate axis. The figure shows an abrupt increases of C at $\alpha \simeq 1$. The coefficient C reaches values between $C = 0.1$ and $C = 0.4$ for $1 < \alpha < 2$, which corresponds to networks dropping as indicated by Eq. 5.3, but with a degree distribution very close to a power law (figure 5.2); for $0 < \alpha < 1$, which corresponds exactly to scale-free networks, C reaches values orders of magnitude larger than those corresponding to the original Barabási-Albert construction. Note that these values for C are of the same order than the ones found in real networks. On the other hand, the study of the coefficient $C(k)$ as function of the degree k also shows an interesting behavior. Panel (b) shows that $C(k)$ decreases as a power law with k, as it has been observed in many real networks. The curves of the picture correspond, from bottom to top, to networks which were generated using

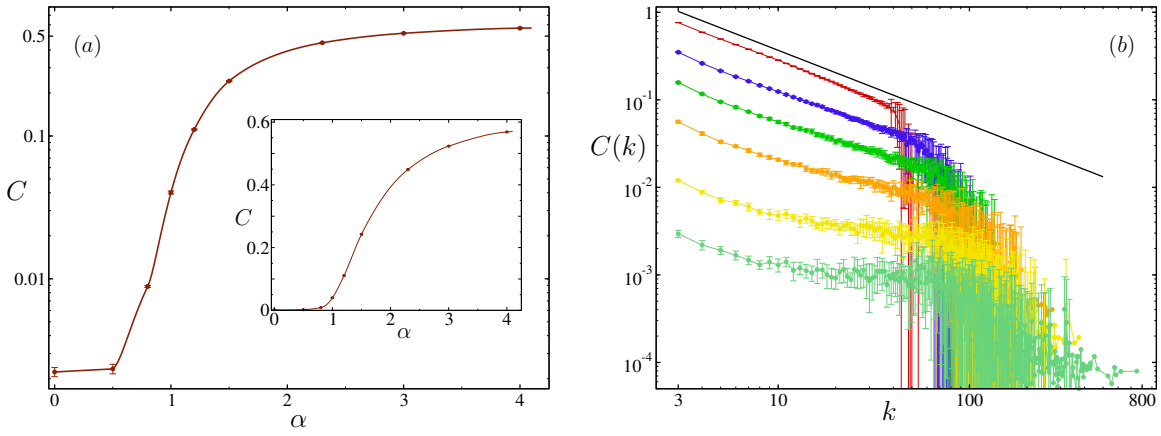

Figure 5.6: Panel (a), the mean clustering coefficient C as function of the model's parameter α. Panel (b), $C(k)$ as function of k for different values of α: $\alpha = 0$, $\alpha = 0.8$, $\alpha = 1.0$, $\alpha = 1.2$, $\alpha = 1.5$, and $\alpha = 3$, from bottom to top. The slope of the upper straight line is -0.85.

$\alpha = 0$ (corresponding to the Barabási-Albert model), $\alpha = 0.8$, $\alpha = 1.0$, $\alpha = 1.2$, $\alpha = 1.5$, and $\alpha = 3$, respectively. The upper straight, which was plotted as a visual guide, has a slope of -0.85.

5.2 Interdependence between geography and vertex attractiveness

The preceding model is possibly the simplest way to combine the two opposing generating mechanisms which, presumably, govern the development of geographical networks: the tendency of vertices to connect to the more "attractive" vertices of the network and the fact that long-range connections are usually penalized due to high cost. The model, in spite of successfully reproducing many properties of real networked systems is, however, still unrealistic in some aspects. For instance, the model requires implicitly to know at each time the global state of the system; to attach a new vertex to the network, one needs to know the location and degree of all vertices already present. In real systems, however, when a new "member" tries to connect in, it does not usually have an exact knowledge of the properties of all other members, rather of most of the members that exist in its physical neighborhood. On the other hand, the model does not take into account the possibility that a connection is established between vertices already present in the network, which often happens in many real systems. Thus, the model can evidently be improved.

We propose in this section another spatial model which accounts for these deficiencies. The basic features that we will desire for this new model are the following: First of all, that the organizing mechanisms under which the network evolves are local. For this, the knowledge about the network of any given vertex must be limited to a certain (Euclidean) neighborhood of the vertex; each vertex is aware of all properties of the vertices belonging to its neighborhood, but nothing of the properties of the rest of the vertices. The range of the neighborhood must be governed by a cost function which establishes the importance of the geographical constraints; as the connection cost grows, the range of the neighborhood must decrease. On the other hand, the network will as usual grow by adding vertices and edges. At each time step, new

vertices will be connected to the network, and, additionally, new edges will be added to the network connecting old vertices. The preferential attachment condition will be implemented by the inclination of vertices to connect to vertices of large degree -the more attractive ones within their neighborhood- lying within their neighborhood. Apart from these requirements (which some authors have already considered [Barthélemy, 2003]), we demand two new ingredients: The interaction cost governing the range of each neighborhood must depend on the attractiveness of the vertex associated; the larger the vertex attractiveness the larger its interaction range must be. The probability that a new vertex appears in an isolated area, geographically far from the rest of the vertices, must be smaller than the probability that the new vertex appears close to the already existing vertices.

The last two conditions are inspired by the properties of the technological networks, in which the cost of establishing long-range connections between distant spots is usually very high. Technological networks are man-made networks designed for transport or distribution of a commodity or resource, such as electricity, wares, persons, or information. Transport links spots distributed over a certain geographical area, or sometimes all over the world. Consider for example the Internet or the airline network. The statistical study of these networks suggests that when long-range connections exist, they usually connect hubs [Yook et al., 2002, Gastner and Newman, 2004]: For instance, transcontinental flights usually link important airports, to which end many airline routes, and long-range connections in the Internet can be found between big telecommunication centers only. This is not surprising. When telecommunication or airline companies decide to make big investments creating long-range transport channels, they want to link important spots, so that the large amount of information, wares, or persons (which will presumably be exchanged between these spots) compensates the expense. On the other hand, the condition that vertices tend to arise close to other already existing vertices, tries to model the fact that, in systems where the long-range interaction cost is high, vertices which are distant from any other vertex do not usually connect to the existing network, i. e., they remain isolated[1]. Consider for example the electric power grids, which are networks where distant connections are enormously high priced. When a power station is constructed in a region too far from "civilization", the station supplies electricity to the buildings close to it, and no high-voltage transmission lines link the station to the distant grid of the "civilized" world; the station then remains isolated. In fact, large electric stations are usually not constructed far from "civilization"; the inhabitants of an isolated region will probably use small generators for personal use, and a station will only be constructed when "civilization" comes to the region. Consider now a network for which the long-range connection cost in not extremely high, for instance the Internet. Routers usually concentrate in towns, rather than in deserted areas, because people live and work in towns; Consequently, there is a larger probability that a new Internet access appears in a town, and therefore, in the vicinity of other already existing accesses. Observe furthermore, that the more industrialized the town, the more rapidly the number of Internet accesses grows. One must, however, also take into account, that vertices do not appear extremely close to each other. Thus, constructing two big power stations in close vicinity to one another (for example, one kilometer apart) is not reasonable; it is cheaper to build only one bigger station which supplies electricity to the entire area. Or, for example, it is not common for a family house to have two routers. Therefore, we assume that there are certain areas, not too far but also not too close to existing vertices, where new vertices will probably arise.

There are many ways of implementing the above conditions. Consider attractiveness, for instance: It must be decided what property of vertices is suitable to characterize the vertex attractiveness; additionally, the mechanism of attachment can depend linearly on the attrac-

[1]This is true, at least, as long as the network is not extended until it has reached the vertex.

tiveness, as happens in the Barabási-Albert construction, or follow another type of dependence. On the other hand, the interaction range of the neighborhoods must decay depending on the distance and attractiveness; but the dependence can be linear, logarithmic, or follow any other mathematical function. Moreover, the probability function which governs the appearance of new vertices in a given point of space can be a complex function involving the positions of all already existing vertices present in the whole network at the time, or depending only on those vertices closer to the point. In any case, extensive numerical simulations show that different (but similarly oriented) prescriptions produce models which differ only slightly. This supports the general value of the conditions proposed. The appropriate prescriptions for a determined pattern should of course be implemented depending on the particular geographical system that we attempt to model. Since we are not interested in studying any particular real-world network, but in capturing the general features of spatial systems, we adopt here very simple prescriptions to implement the required conditions. The soundness of the chosen prescriptions is supported by the fact that they very successfully reproduce the general properties found in spatial networks.

We start from the two-dimensional Euclidean plane. In this space we place at random m_o vertices. The distance between any two of these initial vertices should be larger than a certain r_{min}, the minimum distance that separates vertices in the network. The order of magnitude of the distances between them can be somewhat controlled restricting the positioning of the vertices to a certain preselected area of the Euclidean plane. Around these initial vertices we will let our network grow. Thus, at every time step, we first add a new vertex which attaches through m_1 ($m_1 < m_o$) edges to m_1 different vertices of the network; additionally, once the new vertex is connected, m_2 edges are distributed, each connecting two different vertices of the network (the new one included), that is, if no edge connecting them already exists. In both cases, the edges are added to the network only if the constraints due to the geography allow the addition. If a vertex cannot connect to the other vertices belonging to its neighborhood -because no vertex in the neighborhood exists, or, if they exist, they are already connected to the vertex-, no edge is added. Each new vertex is placed as follows: An old vertex of the preexisting network is primarily chosen at random. It will define an area in the shape of a ring, where the new vertex may be situated. The area results from the intersection of two discs which are centered where the old vertex is positioned, one disc of radius r_{max} and another of radius r_{min} ($r_{max} > r_{min}$). Secondly, a point belonging to this ring-area neighboring the old vertex is randomly picked. Using polar coordinates, we choose at random a radius belonging to the interval $[r_{min}, r_{max}]$ and an angle $\phi \in]0, 2\Pi]$. If the point is a distance smaller than r_{min} from every vertex already present in the network, then the selection is rejected and a different old vertex is chosen; in the opposite case, the new vertex is placed in the selected point. Thus, r_{max} plays the role of the interaction range of every new vertex, and r_{min} the minimum distance between in the network. Note that, when $r_{max} \to \infty$ and $r_{min} \to 0$, the process does not give an homogeneous distribution in space, but essentially means that smaller distances to the chosen vertex are preferred. With respect to vertex attractiveness, we again consider that it is simply characterized by the vertex degree. Thus, the interaction range of vertices will depend directly on the vertex degree. We impose that the function governing the interaction range of each vertex i of degree k_i verifies

$$r_i = r_{max} + \beta {k_i}^{\gamma} \ , \tag{5.4}$$

where β and γ are non-negative tuning parameters whose function is to control the cost associated to the connections. Note that the effects of the geography disappears when $r_{max} \to \infty$. The interaction range r_i determines the range of the physical neighborhood of vertex i, which contains the vertices to which vertex i can connect. Thus, when a new vertex n tries to connect, it "sees" only those vertices that belong to its neighborhood of range $r_n = r_{max}$. The prefer-

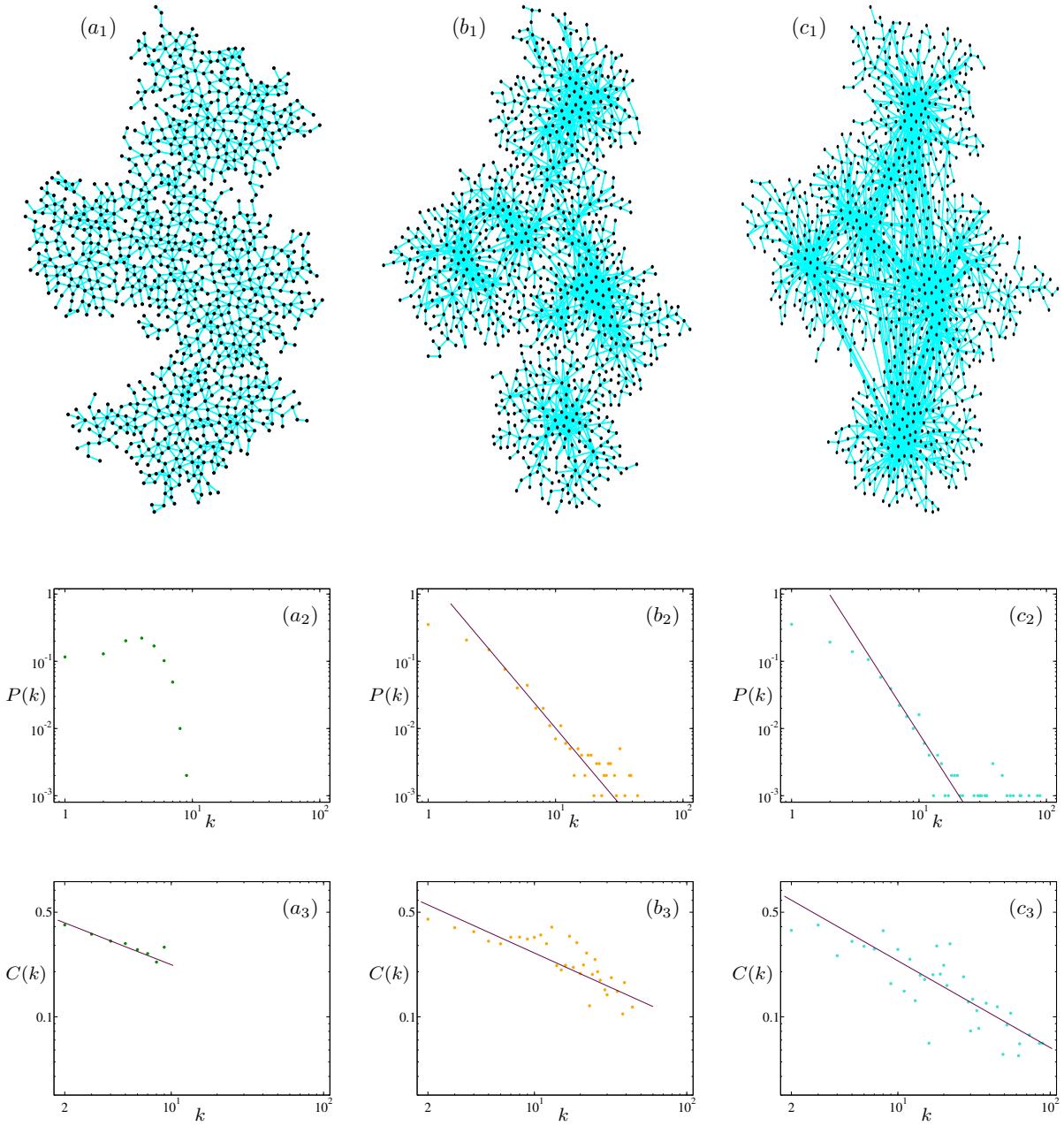

Figure 5.7: (a_1), (b_1), and (c_1): graphical representations of the models (a), (b), and (c), respectively. The three networks have 1000 vertices; their size is $L = 1881$ (a_1), $L = 1982$ (a_2), and $L = 1982$ (a_3). Note that all edges of network (a_1) are short-range connections, while in network (c_1) edges connecting distant vertices do exist. (a_2), (b_2), and (c_2): Degree distribution of the networks represented in (a_1), (a_2), and (a_3), respectively. Models (b) and (c) are scale-free. The slope of the straight lines are -2.24 (b_2) and -2.95 (c_2). (a_3), (b_3), and (c_3): Local clustering coefficient $C(k)$ of the networks represented in (a_1), (a_2), and (a_3), respectively. The behavior of $C(k)$ follows power laws for three models. The slope of the straight lines are -0.40 (a_3), -0.46 (b_3) and -0.58 (b_3). Notice the double logarithmic scales in all graphs.

ential attachment is implemented in the following way: each vertex i connects to the vertex of larger degree belonging to its neighborhood which is not yet connected to vertex i. Finally, the additional process of adding m_2 edges between old vertices works as follows: first, one vertex j of the network is chosen at random, and then it is connected to m_2 vertices that belong to its physical neighborhood following the preferential attachment prescription described; in this case the range d_j of the neighborhood is given by Eq. 5.4.

In the present study we restrict ourselves to two sets of values for the parameters of the model. The first one, which includes three different spatial cases, illustrates the impact of the cost-distance dichotomy on network structure. We consider the following values: $m_1 = 1$, $m_2 = 1$, $r_{min} = 500$ m. u., and $r_{max} = 1000$ m. u., where m. u. stands for the metrical units. The fact that we impose $r_{max} = 2 \cdot r_{min}$, i. e., that r_{max} is only twice r_{min}, indicates that we will deal in this case with networks for which the cost of new vertices for establishing long-range connections is very high. Additionally, we choose $m_0 = 7$, and a radius of 14000 m. u. for the disc-area within which the $m_0 = 7$ initial vertices may be placed at random. We distinguish three different cases: Case (a), $\beta = 1$ and $\gamma = 1.4$ (corresponding to a spatial network in which the geographical constraints are extremely important since long-range connections are practically impossible); Case (b), $\beta = 1.5$ and $\gamma = 2.3$ (an intermediate case); Case (c), $\beta = 2$ and $\gamma = 4$ (for which it is already allowed that vertices of large degree establish long-range connections). Of course, the selected values are somehow arbitrary; we consider these particular values because the emerging networks present similar properties to those found in real networks. Thus, for the case $\beta = 0$ the resulting network is practically a tree, because no edges can be placed between old vertices. On the other hand, for very large β and γ a "winner-takes-all" phenomenon emerges, in which almost all vertices are connected to one super-hub that has an enormous degree.

Figure 5.7 compares the results of simulations corresponding to these three cases. With the purpose of depicting the resulting networks, we consider small networks with only 1000 vertices. The numerical simulations indicate that the topological properties do not significantly change as the order of the networks grows. The graphical representations a_1, b_1, and c_1 of the resulting networks show clearly the effects of the selective growth of the interaction range as a function of the degree of vertices: For systems where long-range connections are extremely expensive (model a), even the more important vertices of the network connect to the few neighbors which are very close to it. As the cost of establishing long-range interactions decreases, connections between distant vertices in the network begin to appear, in particular, between high degree vertices (models b and c). The degree distribution evidently changes as the geographical constraints are gradually loosened. Thus, model a shows a degree distribution which decays approximately as an exponential (panel a_2); no vertices of large degree can be found. The degree distributions of models b and c, however, exhibit well-defined power law tails (in spite of the small order of the networks considered): $P(k) \sim k^{-2.25}$ (panel b_2) and $P(k) \sim k^{-2.95}$ (panel c_2). The local clustering coefficients $C(k)$ also show a behavior very close to that found in real networks (see panels a_3, b_3, and c_3). All three models exhibit power law behaviors for $C(k)$: $C(k) \sim k^{-0.40}$ (panel a_3), $C(k) \sim k^{-0.46}$ (panel b_3), and $C(k) \sim k^{-0.58}$ (panel c_3). Note the double logarithmic scales. In addition, the mean clustering coefficient C of these spatial models is always quite large, about $C \simeq 0.33$ for all of them. The number of triangles found in the networks was, however, 668 (model a), 1593 (model b), and 1341 (model c). On the other hand, the average path length decreases as the amount of long-range connections grows: $l = 20.18$ (model a), $l = 8.83$ (model b), and $l = 5.01$ (model c). This result is natural, and shows the transition from a quasi-planar graph endowed with a structure quite similar to a lattice (model a) to the typical network structure found in most real networks (models b and c).

Because the fact that model a is practically a planar graph, that is, both vertices and edges

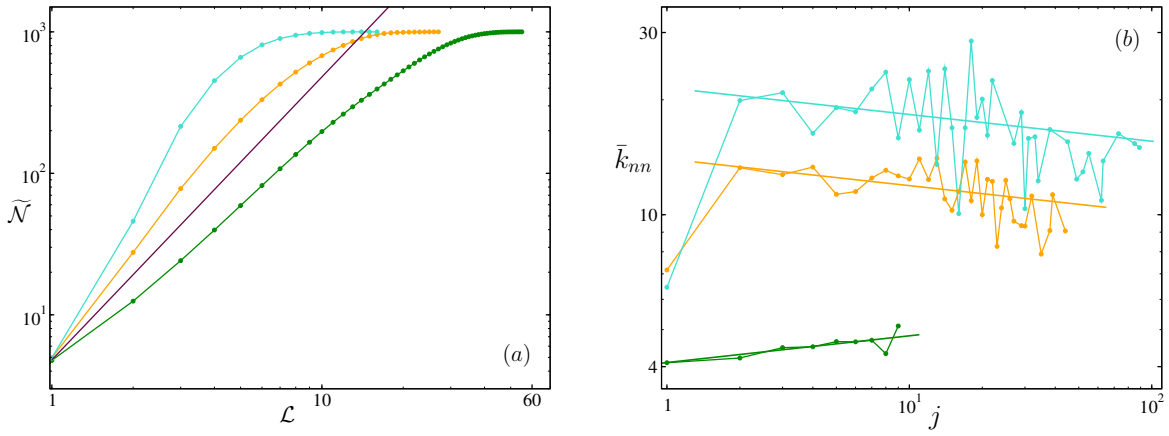

Figure 5.8: (a): Accumulative distribution of the average number of vertices per shell, $\bar{\mathcal{N}}(\mathcal{L})$, as a function of the shell number $\mathcal{L}$. Notice the double logarithmic scales of the picture. The results are used to estimate the dimension of the spatial networks represented in figure 5.7. Model (a), green curve; model (b), orange curve; model (c), cyan curve. The black straight line correspond to a network of dimension 2 (see text for more details). (b): Analysis of the degree-degree correlations of the three models considered. Nearest neighbors' average function $\bar{k}_{nn}(j)$ as a function of j. Green curve, model (a); orange curve, model (b); cyan curve, model (c). Note that models (b) and (c) present dissortative mixing.

are embedded in the two-dimensional space, we expect that $\bar{\mathcal{N}}(\mathcal{L})$, the average number of vertices which can be found with a distance of $\mathcal{L}$ steps or less from a vertex, approximately varies as the square of $\mathcal{L}$. This idea can be used to define an effective "dimension" of networks [Gastner and Newman, 2004]. Figure 5.8 (a) shows on double logarithmic scales how $\bar{\mathcal{N}}(\mathcal{L})$ behaves as a function of $\mathcal{L}$. The black straight line corresponds to the critical behavior $\bar{\mathcal{N}}(\mathcal{L}) \sim \mathcal{L}^2$; the green, orange and cyan curves correspond, respectively, to the behavior of models (a), (b), and (c). We can see that the "dimension" of model a is smaller than two. However, the "dimension" of models b and c -which are obviously not planar graphs- is larger than two.

Figure 5.8 (b) exhibits the correlation properties of these spatial models. In the picture we plot the nearest neighbors' average function, $\bar{k}_{nn}(j)$, as a function of j. The lowest curve, corresponding to model (a), shows that the network is slightly assortative. This feature of model (a) is probably due to the characteristic distribution of vertices in the space. Note that figure 5.7 (a_1) demonstrates that a given vertex usually has the same degree as its neighbors[2]. On the other hand, models (b) and (c) present dissortative mixing; interestingly, for both models $\bar{k}_{nn}(j)$ falls with j following power laws of the form $\bar{k}_{nn}(j) \sim j^{-\iota}$, just like in real networks.

The study of the tomography also reveals interesting characteristics. As we expected, the accumulative distribution of the average number of vertices per shell, $\bar{\mathcal{N}}(\mathcal{L})$, shows that both the diameter and the mean path length increase as geographical effects become more important (figure 5.9 b). More interesting is, however, the peak appearing in panel (a) of the picture, where we plot the average degree $\langle k \rangle_{\mathcal{L}}$ as a function of layer number $\mathcal{L}$. It shows that the mean degree $\langle k \rangle_{\mathcal{L}=1}$ increases rapidly as the number of long-range connections grows in the network; on the other hand, the average shell degree decreases fast for more distant layers, $\mathcal{L} > 1$. This indicates that vertices with large degrees are rapidly found in this type of networks, which has especial

[2]In addition, the areas containing a large density of vertices usually contain a large density of edges, and vice versa, the areas containing a small density of vertices also contain a small density of edges.

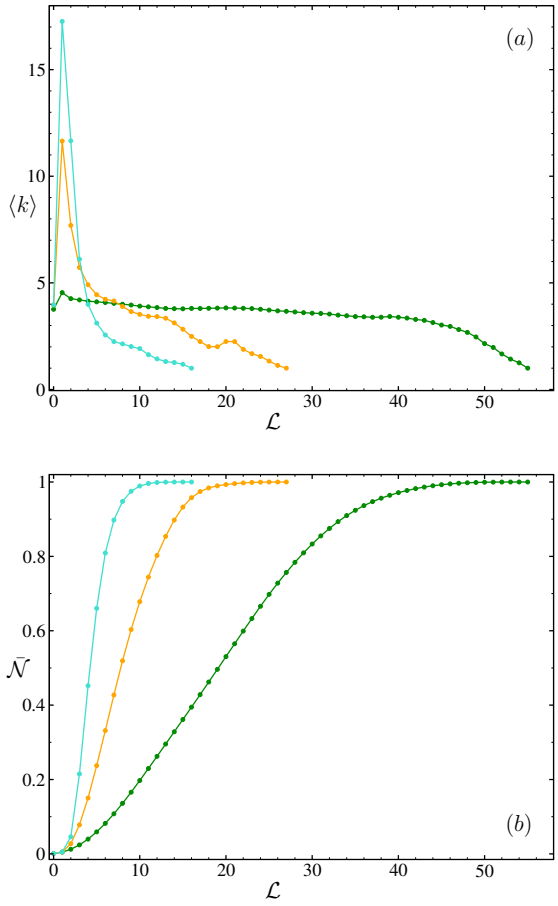

Figure 5.9: (a): Average degree, $\langle k\rangle_{\mathcal{L}}$, as a function of shell number $\mathcal{L}$. As the number of hubs and long-range connections increases, the peak at $\langle k\rangle_{\mathcal{L}=1}$ grows. Note that this result has important effects on epidemiological properties: vertices of large degree are rapidly affected by the spreading of an infection. (b): Accumulative distribution of the average number of vertices per shell, $\bar{\mathcal{N}}(\mathcal{L}) = (\sum_{r,k} N_{\mathcal{L},r}(k))/N$ as function of shell number $\mathcal{L}$. For both pictures the curves correspond to: model (a), green; model (b) orange; model (c), cyan.

importance when dealing with spreading in networks. Evidently, in a network like that of model (a) the propagation of any spreading agent will be similar to the propagation on a lattice: the spreading agent will primarily reach the nearest physical neighbors.

Model (a) quite well reproduces the properties of those systems where vertices and edges are embedded in the two-dimensional physical space, like, for example, electric power grids or networks or roads. However, none of the preceding three models is suitable for characterizing world-scale systems such as the Internet or the network of airline routes. The reason being that, in these systems, vertices are not uniformly distributed in the region under study (more or less like in our preceding models, see pictures 5.7 a_1, b_1, and c_1), but they are usually concentrated in the more technological areas of the world. Thus, a more realistic model for describing such systems must take into account that, both in the Internet and the airline route networks, there are usually many "desert" regions lying between the areas where vertices are found in abundance. Such a pattern is easy to construct by varying the ratio between r_{max} and r_{min} of our model. This aspect is actually considered by our second selection of parameters. As we will see next, inhomogeneous distribution of vertices in space influences quantitatively the statistical properties of networks.

Let us thus consider that the ratio between r_{max} and r_{min} is larger, for example, $r_{max} = 5 \cdot r_{min}$[3]. The values of the parameters for this model (model d) are now the following: $m_0 = 7$,

[3]Note that the change in the ratio modifies not only the distribution of vertices in space, but also makes cheaper the cost for establishing connections of a new vertex; the neighborhood of a new vertex will in comparison contain more vertices.

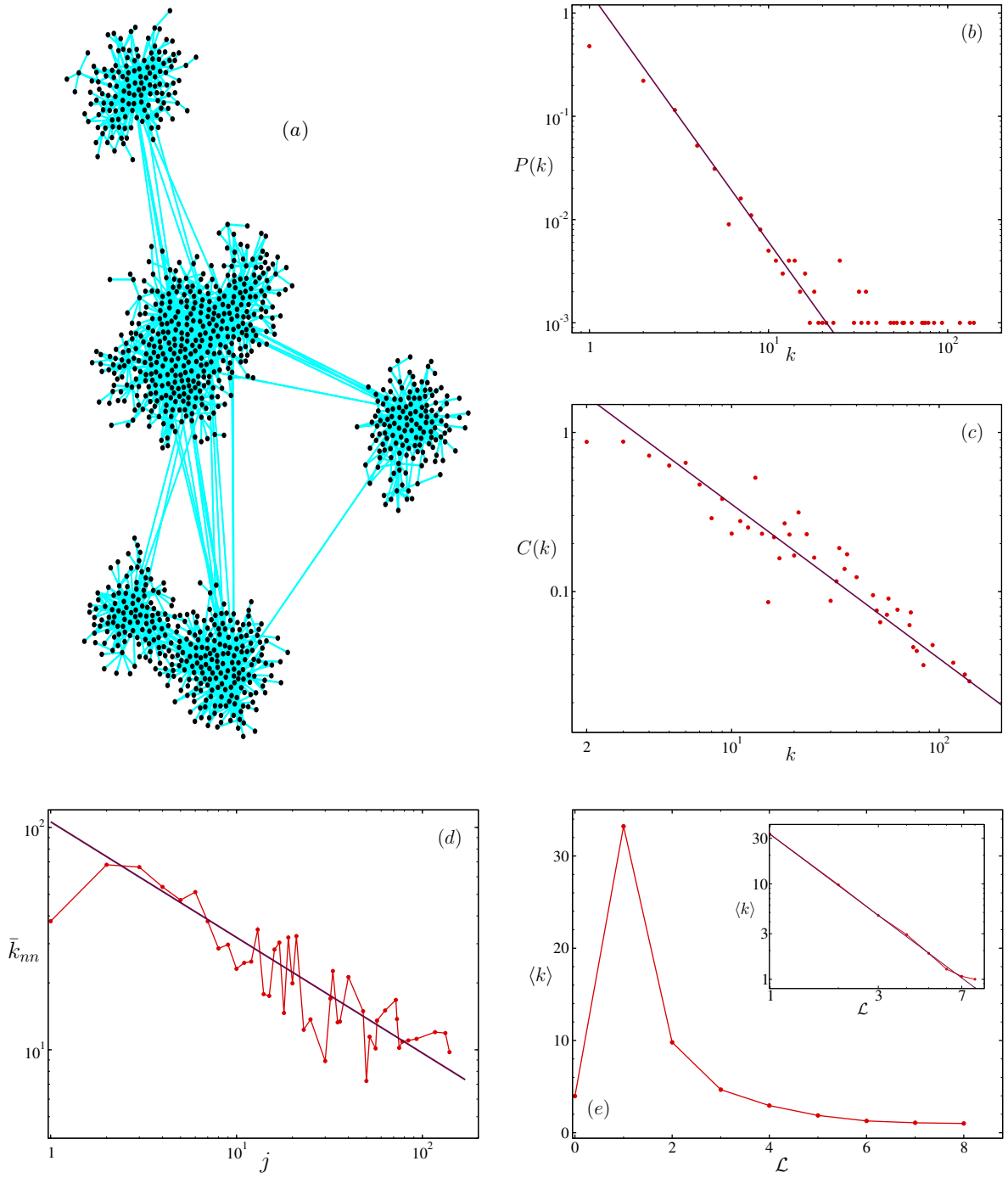

Figure 5.10: (a): Graphical representation of a small network ($N = 1000$ and $L = 1985$) corresponding to model (d). Note that vertices concentrate in certain areas of the space, and that the long-range connections of the network link end-vertices of large degree. (b): Degree distribution of model (d). (c): Degree-dependent clustering coefficient $C(k)$ of the model. (d): nearest neighbors' average function $\bar{k}_{nn}$ as function of the degree j. Note that $P(k)$, $C(k)$, and $\bar{k}_{nn}$ fall off as power law functions. (e): Average degree, $\langle k \rangle_{\mathcal{L}}$, as a function of shell number $\mathcal{L}$. From $\langle k \rangle_{\mathcal{L}=1}$ the average degree decays interestingly following a power law (inset of the picture).

$m_1 = 1$, $m_2 = 1$, $r_{min} = 200$, $r_{max} = 1000$, $\beta = 2$ and $\gamma = 3$. As before, the initial area, where the $m_0 = 7$ vertices are randomly placed, has a radius of 14000 m. u. and the order of the network is $N = 1000$. Figure 5.10 (a) shows this network in two-dimensional Euclidean space: The model simulates perfectly the tendency of vertices to concentrate in different areas having a high density of vertices (as if these areas were urban centers, i. e., cities or a conglomeration of cities), which are interconnected through long-range connections which always join vertices of large degree (usually belonging to different geographical communities).

The properties of this simple model are especially interesting, since they reproduce quite well the structural properties found in the Internet. Thus, the degree distribution of the model follows a power law function of the form: $P(k) \sim k^{-2.42}$ (panel b of figure 5.10). The mean clustering coefficient is large. For our small network $C \simeq 0.7$; for realizations of the model dealing with networks of order N larger, C decreases slowly. The degree-dependent clustering coefficient $C(k)$ decays as power law function $C(k) \sim k^{-0.97}$ (picture 5.10 c). The decay of $C(k)$ is, however, not as pronounced when N is larger. The average path length is very small, $l = 3.74$, and numerical simulations with larger networks indicate the existence of small-world behavior. In addition, the network shows dissortative mixing: the nearest neighbors' average function $\bar{k}_{nn}$ decreases as $\bar{k}_{nn}(j) \sim j^{-0.52}$ (see picture 5.10 d). Let us now remember that i) the degree distribution of the Internet follows a power law with exponent $\gamma \simeq -2.5$, ii) the local clustering function $C(k)$ behaves as $C(k) \sim k^{-0.75}$, and iii) $\bar{k}_{nn}$ decreases with j following the function $\bar{k}_{nn}(j) \sim j^{-0.5}$ [Vázquez et al., 2002]. The agreement with the Internet is therefore excellent. Finally, we plot in figure 5.10 (e) the average degree $\langle k \rangle_{\mathcal{L}}$ as a function of shell number $\mathcal{L}$ corresponding to the study of the tomography of the model. We see again that hubs are found only a few steps away from any vertex, and interestingly, that $\langle k \rangle_{\mathcal{L}}$ drops as a perfect power law from $\mathcal{L} = 1$ on (see inset of the picture; double logarithmic scales).

To conclude, we emphasize the fact that the described model is only one of many possible models we could have devised according to the guidelines and conditions proposed at the beginning of this section. In fact, we are convinced that other prescriptions could possibly reproduce the properties of (certain) real networks in an even more successful way[4]. However, the results that our very simple model yields seem to suggest that the basic mechanism proposed, the dependence of the geographical constraints on the attractiveness of the elements of the system, plays an important role for the development and topology of spatial networks.

[4]Bear in mind that the ingredients that must be introduced in a model in order to reproduce the properties of a real system must naturally be inspired by the particular characteristics and working of the system under study.

Chapter 6

Conclusions

In this thesis we review some important topics in the field of networks and introduce some new results which, we hope, provide new insight into the intrinsic properties and topology of these relevant structures.

We explore the effects of degree-degree correlations and transitivity in the structure of scale-free networks. In order to carry out this survey, we introduce three new algorithms which change the degree-degree correlations and transitivity of networks in such a way that assortative, dissortative, and transitive mixing (the most outstanding types of vertex-pair correlations found in the real-world systems) are respectively produced to a desired degree. All three algorithms are based on link-restructuring processes, which does not change the degree sequence of the networks to which they are applied. Making use of these algorithms, we demonstrate that assortativity, dissortativity, and transitivity influence considerably the average path length and mean clustering coefficient of networks, as well as their tomographical properties. The importance of these results should not be underrated, since they indicate that vertex-pair correlations are essential for making a correct description of the spreading phenomena (such as the spreading of information or infections) taking place on networks.

As a second step, we use these algorithms to investigate the percolation problem on correlated scale-free networks. We show by means of extensive numerical simulations that the structure of strongly assortative networks is severely damaged when removing a small fraction of vertices or edges; however, a very robust "core"-component seems to remain eternally in the system, irrespective of the mass of elements eliminated (even when $\langle k^2 \rangle$ is finite). This implies that highly assortative networks undergo no phase transition. On the other hand, strongly dissortative networks are more robust than those of the uncorrelated type under removal of moderated masses of vertices or edges. However, they do experience a true percolation transition when the fraction of eliminated elements is large enough (even when $\langle k^2 \rangle$ is infinite). The phase transition takes place at amounts of removed elements smaller than in uncorrelated networks. Numerical and theoretical studies of highly triangulated networks show that this type of network undergoes true percolation transitions when the degree of transitivity is large (even when $\langle k^2 \rangle$ is infinite). The phase transitions occur at smaller and smaller fractions of removed elements as the mean clustering coefficient grows.

The three algorithms are mathematical tools which we use to investigate and analyze the topology and percolation properties of networks. However, they are not realistic models for describing physically networked systems. We think, following the ideas of Barabási and Albert, that real networks evolve under few dynamic principles which are responsible for their properties. In relation to this, we are convinced that one fundamental ingredient governing the development of many real networks is, apart from "growth" and "preferential attachment", the high cost of

establishing geographical long-range connections between the different parts of a system. The fact that real networks are embedded in a physical space implies that not all connections are equally probable: neighboring vertices must connect between themselves with larger probability than distant vertices. Two simple models of evolving spatial networks are finally proposed. Both models are based on the competition between preferential attractiveness (edges are preferably attached to attractive vertices) and geographical penalization (long-range edges are disadvantaged compared to short-range edges). The fist model is a one-dimensional extension of the classical Barabási-Albert construction, and the second, a much more realistic model, exploits in a two-dimensional Euclidean space the possible interdependence between geographical cost and vertex-attractiveness. The significance of both models is that they are capable of reproducing many of the properties found in real-world networks. The models generate networks with a high degree of transitivity, small-world behavior, and, depending on the model parameters, either power-law (scale-free character) or exponential tailed degree distributions. Moreover, the networks constructed using the prescription given in the second model, show that the local clustering coefficient and the nearest neighbors' average function drop following power law functions with critical exponents which are very similar to those found in some technological networks. The results suggest that geography plays a key role in the evolution of networks.

Bibliography

L. A. Adamic. *Lecture Notes in Computer Science*, volume 1696 of *Research and Advanced Technology for Digital Libraries*, pages 443–452. Springer Verlag, New York, 1999.

L. A. Adamic and B. A. Huberman. Power-law distribution of the world wide web. *Science*, 287: 2115, 2000.

R. Albert and A.-L. Barabási. Topology of evolving networks: Local events and universality. *Phys. Rev. Lett.*, 85: 5234–5237, 2000.

R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74: 47–96, 2002a.

R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74: 47–97, 2002b.

R. Albert, H. Jeong, and A.-L. Barabási. Internet: Diameter of the world-wide web. *Nature*, 401: 130–131, 1999.

R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406: 378–382, 2000.

L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley. Classes of small-world networks. *Proc. Natl. Acad. Sci.*, 97: 11149–11152, 2000.

C. Andersson, K. Lindgren, S. Rasmussen, and R. White. Urban growth simulation from "first principles". *Phys. Rev. E*, 66: 026204, 2002.

K. Appel and W. Hake. Every planar map is four colorable. part. i. discharging. *Illinois J. Math.*, 21: 429, 1977a.

K. Appel and W. Hake. Every planar map is four colorable. part. ii. reducibility. *Illinois J. Math.*, 21: 491–567, 1977b.

A.-L. Barabási. *Linked: The new Science of Networks*. Perseus, Cambridge, MA, 2002.

A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286: 509–512, 1999.

A.-L. Barabási, R. Albert, and H. Jeong. Mean field theory for scale-free random networks. *Physica A*, 272: 173–187, 1999.

A.-L. Barabási, R. Albert, and H. Jeong. Scale-free characteristics of random networks: The topology of the world wide web. *Physica A*, 281: 69–77, 2000a.

A.-L. Barabási, R. Albert, H. Jeong, and G. Bianconi. Power-law distribution of the world wide web. *Science*, 287: 2115, 2000b.

A.-L. Barabási, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Viseck. Evolution of the social network of scientific collaborations. *Physica A*, 311: 590–614, 2002.

G. I. Barenblatt, editor. *Scaling*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2003.

A. Barrat and R. Pastor-Satorras. Rate equation approach for correlations in growing network models. *Phys. Rev. E*, 71: 036127, 2005.

A. Barrat and M. Weigt. On the properties of small-world networks models. *Eur. Phys. J. B*, 13: 547–560, 2000.

A. Barrat, M. Barthélemy, and A. Vespignani. The effects of spatial constraints on the evolution of weighted complex networks. *J. Stat. Mech.*, P05003, 2005.

M. Barthélemy. Crossover from scale-free to spatial networks. *Europhys. Lett.*, 63: 915–921, 2003.

M. Barthélémy and L. A. N. Amaral. Small-world networks: Evidence for a crossover picture. *Phys. Rev. Lett.*, 82: 3180–3183, 1999a.

M. Barthélémy and L. A. N. Amaral. Erratum: Small-world networks: Evidence for a crossover picture [phys. rev. lett. 82, 3180 (1999)]. *Phys. Rev. Lett.*, 82: 5180, 1999b.

D. ben Avraham, A. F. Rozenfeld, R. Cohen, and S. Havlin. Geographical embedding of scale free networks. *Physica A*, 330: 107, 2003.

E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai, editors. *Complex Networks*, volume 650 of *Lecture Notes in Physics*. Springer-Verlag, 2004.

E. A. Bender and E. R. Canfield. The asymtotic number of labeled graphs with given degree sequences. *J. Comb. Theory, Ser. A*, 24: 296–307, 1978.

J. Berg, M. Lässig, and A. Wagner. Structure and evolution of protein interaction networks: a stistical model for link dynamics and gene duplications. *BMC Evolutionary Biology*, 4: 51, 2004.

A. Bhan, D. J. Galas, and T. G. Dewey. A duplication growth model of gene expresion networks. *Bioinformatics*, 18: 1486–1493, 2002.

G. Bianconi and A.-L. Barabási. Competition and multiscaling in evolving networks. *Europhys. Lett.*, 54: 436–442, 2001a.

G. Bianconi and A.-L. Barabási. Bose-einstein condensation in complex networks. *Phys. Rev. Lett.*, 86: 5632–5635, 2001b.

G. Bianconi and A. Capocci. Number of loops of size h in growing scale-free networks. *Phys. Rev. Lett.*, 90: 078701, 2003.

G. Bianconi, G. Calderelli, and A. Capocci. Loops structure of the internet at the autonomous system level. *Phys. Rev. E*, 71: 066116, 2005.

J. J. Binney, N. J. Dowrick, A. J. Fisher, and M. E. J. Newman. *The Theory of Critical Phenomena: An Introduction to the Renormalization Group*. Oxford University Press, 1992.

M. Boguñá and R. Pastor-Satorras. Class of correlated random networks with hidden variables. *Phys. Rev. E*, 68: 036112, 2003.

M. Boguñá, R. Pastor-Satorras, and A. Vespignani. Absence of epidemic threshold in scale-free networks with degree correlations. *Phys. Rev. Lett.*, 90: 028701, 2003.

M. Boguñá, R. Pastor-Satorras, and A. Vespignani. Cutt-offs and finite size effects in scale-free networks. *Eur. Phys. J. B*, 38: 205–209, 2004.

B. Bollobas. *Random Graphs*. Academic Press, 2nd ed. edition, 2001.

B. Bollobas. *Modern Graph Theory*. Springer-Verlag, 1998.

B. Bollobás. Degree sequences on random graphs. *Discrete Math.*, 33: 1–19, 1981.

B. Bollobás. The diameter of random graphs. *IEEE Trans. Inform. Theory*, 36: 285–288, 1990.

B. Bollobás and O. Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24: 5–34, 2004.

B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, 18: 279–290, 2001.

S. Bornholdt and H. Ebel. World wide web scaling exponent from simon's 1955 model. *Phys. Rev. E*, 64: 035104, 2001.

S. Bornholdt and H. G. Schuster, editors. *Handbook of graphs and Networks*. Wiley-VCH, Berlin, 2003.

M. Brede and S. Sinha. Assortative mixing by degree makes a network more unstable. *cond-mat/0507710*, 2005.

S. R. Broadbent. In discussion of symposium of monte carlo methods. *J. Roy. Stat. Soc. (B)*, 16: 68, 1954.

S. R. Broadbent and J. M. Hammersley. Percolation processes. *Proc. Cambr. Phil. Soc.*, 53: 629–645, 1957.

A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer Networks*, 33: 309–320, 2000.

A. Bunde, S. Havlin, and M. Porto. Are branched polymers in the universality class of percolation? *Phys. Rev. Lett.*, 74: 2714–2716, 1995.

CAIDA. The cooperative association for internet data analysis. http://www.caida.org/, www.

G. Calderelli, A. Capocci, P. de los Rios, and M. A. Mu noz. Scale-free networks from varying vertex intrinsic fitness. *Phys. Rev. Lett.*, 89: 258702, 2002.

D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Network robustness and fragility: Percolation on random graphs. *Phys. Rev. Lett.*, 85: 5468–5471, 2000.

X. Campi, H. Krivine, N. Sator, and E. Plagnol. Analyzing fragmentation of simple fluids with percolation theory. *Eur. Phys. J. D*, 11: 233–238, 2000.

M. Catanzaro, G. Caldarelli, and L. Pietronero. Assortive model for social networks. *Phys. Rev. E*, 70: 037101, 2004.

M. Catanzaro, M. Boguñá, and R. Pastor-Satorras. Generation of uncorrelated random scale-free networks. *Phys. Rev. E*, 71: 027103, 2005.

F. Chung and L. Lu. The diameter of random sparse graphs. *Adv. Appl. Math.*, 26: 257–279, 2001.

F. Chung, L. Lu, T. G. Dewey, and D. J. Galas. Duplication models for biological networks. *Journal of Computational Biology*, 10: 677–687, 2003.

A. Clauset, M. E. J. Newman, and C. Moore. Finding community structure in very large networks. *Phys. Rev. E*, 70: 066111, 2004.

J. E. Cohen. Threshold phenomena in random structures. *Discrete Appl. Math.*, 19: 113–128, 1988.

R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Resilence of the internet to random breakdowns. *Phys. Rev. Lett.*, 85: 4626–2628, 2000a.

R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Breakdown of the internet under intentional attack. *Phys. Rev. Lett.*, 86: 3682–3685, 2000b.

R. Cohen, , D. ben Avraham, and S. Havlin. Percolation critical exponents in scale-free networks. *Phys. Rev. E*, 66: 036113, 2002a.

R. Cohen, D. Dolev, S. Havlin, T. Kalisky, O. Mokryn, and Y. Shavitt. Leibniz center technical report 2002-49. At http://leibniz.cs.huji.ac.il/research, 2002b.

H. Daido and K. Nakanishi. Aging transition and universal scaling in oscillator networks. *Phys. Rev. Lett.*, 93: 0104101, 2004.

J. Dall and M. Christensen. Random geometric graphs. *Phys. Rev. E*, 66: 016121, 2002.

G. F. Davis and H. R. Greve. Corporate elite networks and governance changes in the 1980s. *Amer. J. Sociol.*, 103: 1–37, 1997.

L. de Arcangelis, S. Redner, and A. Coniglio. Anomalous voltage distribution of random resistor networks and a new model for the backbone at the percolation threshold. *Phys. Rev. B*, 31: 4725–4727, 1985.

P. G. de Gennes and E. Guyon. Lois générales pour l'injection d'un fluide dans un milieu poreux aléatoire. *J. Mec.*, 17: 403, 1978.

M. Argollo de Menezes, C. F. Moukarzel, and T. J. P. Penna. First-order transition in small-world networks. *Europhys. Lett.*, 50: 574–579, 2000.

D. J. de S. Price. Networks of scientific papers. *Science*, 149: 510–515, 1965.

D. J. de S. Price. A general theory of bibliometric and other cumulative advantage processes. *J. Amer. Soc. Inform. Sci.*, 27: 292–306, 1976.

M. di Bernardo, F. Garofalo, and F. Sorrentino. Synchronizability of degree correlated networks. *cond-mat/0504335*, 2005.

DIP. Database of interacting proteins. http://dip.doe-mbi.ucla.edu/, www.

D. N. Dorogovtsev and J. F. F. Mendes. Scaling properties of scale-free evolving networks: Continous approach. *Phys. Rev. E*, 63: 056125, 2001a.

S. N. Dorogovtsev. Clustering of correlated networks. *Phys. Rev. E*, 69: 027104, 2004.

S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. *Adv. in Phys.*, 51: 1079–1187, 2002.

S. N. Dorogovtsev and J. F. F. Mendes. *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, Oxford, 2003a.

S. N. Dorogovtsev and J. F. F. Mendes. Effect of the accelerating growth of communications networks on their structure. *Phys. Rev. E*, 63: 025101, 2001b.

S. N. Dorogovtsev and J. F. F. Mendes. *Accelerated growth of networks*, pages 318–341. in Handbook of Graphs and Networks, S. Bornholdt and H. G. Schuster, eds., Wiley-WCH, Berlin, 2003b.

S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks with aging of sites. *Phys. Rev. E*, 62: 1842–1845, 2000a.

S. N. Dorogovtsev and J. F. F. Mendes. Scaling behavior of developing and decaying networks. *Europhys. Lett.*, 52: 33–39, 2000b.

S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Structure of growing networks with preferential linking. *Phys. Rev. Lett.*, 85: 4633–4636, 2000.

J. P. K. Doye. Network topology of a potential energy landscape: A static scale-free network. *Phys. Rev. Lett.*, 88: 238701, 2002.

J. P. K. Doye and Claire P. Massen. Self-similar disk packings as model spatial scale-free networks. *Phys. Rev. E*, 71: 016128, 2005a.

J. P. K. Doye and Claire P. Massen. Characterizing the network topology of the energy landscapes of atomic clusters. *J. Chem. Phys.*, 122: 084105, 2005b.

V. M. Eguíluz and K. Klemm. Epidemic threshold in structured scale-free networks. *Phys. Rev. Lett.*, 89: 108701, 2002.

M. Eigen and P. Schuster. *The Hypercycle: A Principle of Natural Self-Organization*. Springer-Verlag, NY, 1979.

P. Erdös and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5: 17–61, 1960.

P. Erdös and A. Réyi. On random graphs. *Publ. Math. Debrecen*, 6: 290–297, 1959.

P. Erdös and A. Réyi. On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hungar.*, 12: 261–267, 1961.

G. Ergün and G. J. Rodgers. Growing random networks with fitness. *Physica A*, 303: 261–272, 2002.

J. W. Essam. Percolation theory. *Rep. Prog. Phys.*, 43: 833, 1980.

J. W. Essam. *Percolation and cluster size.* pp. 197-270 in Phase transitions and critical phenomena, Vol II, Domb, C. and Green M. S., academic press edition, 1972.

L. Euler. Solutio problematis ad geometrian situs pertinentis. *Comm. Acad. Sci. Imp. Petrop.*, 8: 128–140, 1736.

M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-paw relationships or the internet topology. *Comput. Commun. Rev.*, 29: 251, 1999.

I. Farkas, I. Derényi, G. Palla, and T. Vicsek. Equilibrium statistical mechanics of network structrures. *Lecture Notes in Physics, Springer-Verlag*, 650: 163–187, 2004.

I. J. Farkas, I. Derényi, A.-L. Barabási, and T. Vicsek. Spectra of "real-world" graphs: Beyond the semicircle law. *Phys. Rev. E*, 64: 026704, 2001.

D. Fell and A. Wagner. The small world of metabolism. *Nature Biotechnology*, 18: 1121, 2000.

G. W. Flake, S. R. Lawrence, C. L. Giles, and F. M. Goetzee. Self-organization and identification of web communities. *IEEE Computer*, 35: 66–71, 2002.

B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, Cambridge, 1985.

S. Fortunato, V. Latora, and M. Marchiori. Method to find community structures based on information centrality. *Phys. Rev. E*, 70: 056104, 2004.

M. J. Gagen and J. S. Mattick. Accelerating, hyperaccelerating, and decelerating networks. *Phys. Rev. E*, 72: 016123, 2005.

M. T. Gastner and M. E. J. Newman. The spatial structure of networks. *arXiv:cond-mat/0407680*, 2004.

K.-I. Goh, B. Kahng, and D. Kim. Spectra and eigenvectors of scale-free networks. *Phys. Rev. E*, 64: 051903, 2001.

J. Gómez-Gardeñes and Y. Moreno. Local versus global knowledge in the barabási-albert scale-free network model. *Phys. Rev. E*, 69: 037103, 2004.

K. B. Hajra and P. Sen. Phase transitions in an aging network. *Phys. Rev. E*, 70: 056103, 2004.

W. R. Hamilton. *Letter to John T. Graves on the Icosian, 19 Oct. 1856.* In:(H. Halberstam and R. E. Ingram, eds.), The mathematica papers of Sir William Rowan Hamilton, Vol. (Algebra). Cambridge University Press, 1931.

J. M. Hammersley. Comparison of atom and bond percolation. *J. Math. Phys.*, 2: 728–733, 1961.

J. M. Hammersley. Percolation processes. lower bounds for the critical probability. *Ann. Math. Statist.*, 28: 790–795, 1957.

J. M. Hammersley and D. J. A. Welsh. Percolation theory and its ramifications. *Contemp. Phys.*, 21: 593–605, 1980.

F. Harary. *Graph Theory.* Perseus, Cambridge, 1995.

T. E. Harris. A lower bound for the critical probability in a certain percolation process. *Proc. Cambr. Phil. Soc.*, 56: 13–20, 1960.

B. Hayes. Graph theory in practice: Part i. *Amer. Sci.*, 88: 9–13, 2000a.

B. Hayes. Graph theory in practice: Part ii. *Amer. Sci.*, 88: 104–109, 2000b.

C. L. Henley. Statics of a "self-organized" percolation model. *Phys. Rev. Lett.*, 71: 2745–2748, 1993.

C. Herrmann, M. Barthélemy, and P. Provero. Connectivity distribution of spatial networks. *Phys. Rev. E*, 68: 026128, 2003.

H.-P. Hsu and M.-C. Huang. Percolation thresholds, critical exponents, and scaling functions on planar random lattices and their duals. *Phys. Rev. E*, 60: 6361–6370, 1999.

L. Huang, L. Yang, and K. Yang. Synchronizability of degree correlated networks. *cond-mat/0503147*, 2005a.

L. Huang, L. Yang, and K. Yang. Enhancing robustness and immunization in geographical networks. *cond-mat/0503147*, 2005b.

L. Hufnagel, D. Brockmann, and T. Geisel. Granular collapse as a percolation transition. *Phys. Rev. E*, 60: 7137, 1999.

L. Hufnagel, D. Brockmann, and T. Geisel. Forecast and control of epidemics in a globalized world. *Proc. Natl. Acad. Sci. USA*, 101: 15124, 2004.

D.-U. Hwang, M. Chavez, A. Amann, and S. Boccaletti. Synchronization in complex networks with age ordering. *Phys. Rev. Lett.*, 94: 0138701, 2005.

M. B. Isichenko. Percolation, statistical topography, and transport in random media. *Rev. Mod. Phys.*, 64: 961, 1992.

T. Ito, K. Tashiko, S. Muta, R. Ozawa, T. Chiba, M. Nishizawa, K. Yamamoto, S. Kuhara, and Y. Sakaki. Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interations in all possible combinations between the yeast proteins. *Proc. Natl. Acad. Sci. U.S. A.*, 97: 1143–1147, 2000.

T. Ito, T. Chiba, , R. Ozawa, M. Hattori, and Y. Sakaki. A comprehensive two-hybrid analysis to explore the yeast protein-protein interactome. *Proc. Natl. Acad. Sci. U.S. A.*, 98: 4569–4574, 2001.

S. Itzkovitz, R. Milo, N. Kashtan, G. Ziv, and U. Alon. Subgraphs in random networks. *Phys. Rev. E*, 68: 026127, 2003.

S. Itzkovitz, R. Milo, N. Kashtan, M. E. J. Newman, and U. Alon. Reply to "comment on 'subgraphs in random networks' ". *Phys. Rev. E*, 70: 058102, 2004.

Jr J. S. Andrade, H. Herrmann, R. F. S. Andrade, and L. R. da Silva. Apollonian networks:simultaneously scale-free, small world, euclidean, space filling, and with matching graphs. *Phys. Rev. Lett.*, 94: 018702, 2005.

H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407: 651, 2000.

H. Jeong, S. P. Mason, Z. N. Oltvai, and A.-L. Barabási. Lethality and centrality in protein networks. *Nature*, 411: 41, 2001.

H. Jeong, Z. Néda, and A.-L. Barabási. Measuring preferential attachment in evolving networks. *Europhys. Lett.*, 61: 567–572, 2003.

J. Joo and J. L. Lebowitz. Behavior of susceptible-infected-susceptible epidemics on heterogeneous networks with saturation. *Phys. Rev. E*, 69: 066105, 2004.

J. Jost and M. P. Joy. Evolving networks with distance preferences. *Phys. Rev. E*, 66: 036126, 2002.

L. P. Kadanoff. The introduction of the idea that exponents could be derived from real-space scaling arguments. *Physics*, 2: 266, 1966.

V. K. Kalapala, V. Sanwalani, and C. Moore. The structure of the united states road network. *preprint, University of New Mexico, Alburquerque*, 2003.

R. Kannan, P. Tetali, and S. Vempala. Simple markov-chain algorithm for generating bipartite graphs and tournaments. *Rand. Struct. Alg.*, 14: 293–308, 1999.

H. Kesten. *Percolation Theory for Mathematicians*. Birkhauser, Boston, 1982.

J. Kim, P. L. Krapivsky, B. Kahng, and S. Redner. Infinite-order percolation and giant fluctuations in a protein interaction network. *Phys. Rev. E*, 66: 055101, 2002.

O. D. King. Comment on "subgraphs in random networks". *Phys. Rev. E*, 70: 058101, 2004.

S. Kirkpatrick. Percolation and conduction. *Rev. Mod. Phys.*, 45: 574, 1973.

V. Klee and D. Larman. Diameters of random graphs. *Can. J. Math.*, 33: 618–640, 1981.

J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. *The web as a graph: Measurements, models, and methods*, volume 1627 of *Lecture Notes in Computer Science*, pages 1–18. Proceedings of the International Conference on Combinatorics and Computing. Springer Verlag, New York, 1999.

P. L. Krapivsky and S. Redner. Organization of growing random networks. *Phys. Rev. E*, 63: 066123, 2001.

P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Phys. Rev. Lett.*, 85: 4629–4632, 2000.

P. L. Krapivsky, G. J. Rodgers, and S. Redner. Degree distributions of growing networks. *Phys. Rev. Lett.*, 86: 5401–5404, 2001.

P. L. Krapivsky, , and S. Redner. A statistical physics perspective on web growth. *Computer Networks*, 39: 261–276, 2002.

L. Kullmann and J. Kertész. Preferential growth: Exact solution of the time-dependent distributions. *Phys. Rev. E*, 63: 051112, 2001.

R. Kumar, P. Raghavan, S. Rajalopagan, and A. Tomkins. Internet: Diameter of the world-wide web. In *Proceedings of the 9th ACM Symposium on Principles of Databases*, page 1, 1999.

R. Kumar, P. Raghavan, S. Rajalopagan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastics models for the web graph. In *Proceedings of the 42th Annual IEEE Symposium on the Foundations of Computer Sciences*, pages 57–65, 2000.

D. Lairez, D. Durand, and J. R. Emery. The chemical gelation viewed through a percolation model simulation. *J. Phys. II*, 1: 977–993, 1991.

V. Latora and M. Marchiori. Is the boston subway a small-world network? *Physica A*, 314: 109–113, 2002.

H. Y. Lee, H. Y. Chan, and P. M. Hui. Scale-free networks with tunable degree-distribution exponents. *Phys. Rev. E*, 69: 067102, 2004.

F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Åberg. The web of the human sexual contacts. *Nature*, 411: 907–908, 2001.

Z. Liu, Y.-C. Lai, and N. Ye. Propagation and immunization of infection on general networks with both homogeneous and heterogeneous components. *Phys. Rev. E*, 67: 031911, 2003.

Lumeta. Internet mapping project. http://research.lumeta.com/ches/map/, www.

J. Machta. Phase transitions in fractal porous media. *Phys. Rev. Lett.*, 66: 169–172, 1991.

H. A. Makse, Jr. J. S. Andrade, M. Batty, S. Havlin, and H. E. Stanley. Modeling urban growth patterns with correlated percolation. *Phys. Rev. E*, 58: 7054–7062, 1998.

S. S. Manna and A. Kabakçioğlu. Scale-free network on euclidean space optimized by rewiring of links. *J. Phys. A: Math. Gen.*, 36: 279–285, 2003.

S. S. Manna and P. Sen. Modulated scale-free network in euclidean space. *Phys. Rev. E*, 66: 066114, 2002.

S. Maslov and K. Sneppen. Specificity and stability in topology of protein networks. *Sience*, 296: 910–913, 2002.

S. Maslov, K. Sneppen, and A. Zalinznyak. Detection of topological patterns in complex networks: Correlation profile of the internet. *Physica A*, 333: 529–540, 2004.

C. P. Massen and J. P. K. Doye. Identifying communities within energy landscapes. *Phys. Rev. E*, 71: 046101, 2005.

N. Masuda, H. Miwa, and N. Konno. Geographical threshold graphs with small-world and scale-free properties. *Phys. Rev. E*, 71: 036108, 2005.

R. M. May and A. L. Lloyd. Infection dynamics on scale-free networks. *Phys. Rev. E*, 64: 066112, 2001.

S. Milgram. The small world problem. *Psych. Today*, 2: 60–67, 1967.

R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Networks motifs: Simple building blocks of complex networks. *Science*, 298: 824–827, 2002.

R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon. On the uniform generation of random graphs with prescribed degree sequences. *arXiv:cond-mat/0312028*, 2004.

M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Rand. Struct. Alg.*, 6: 161–179, 1995.

M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Comb. Prob. Comp.*, 7: 295–305, 1998.

R. Monasson. Diffusion, localization and dispersion relations on "small-world" lattices. *Eur. Phys. J. B*, 12: 555–567, 1999.

K. Moon. Critical behavior of superfluid helium-4 in aerogel. *Phys. Rev. Lett.*, 75: 1328–1331, 1995.

C. Moore and M. E. J. Newman. Exact solution of site and bond percolation on small-world networks. *Phys. Rev. E*, 62: 7059–7064, 2000.

Y. Moreno, R. Pastor-Satorras, and A. Vespignani. Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B*, 26: 521–529, 2002.

M. E. J. Newman. Models of the small world. *J. Statist. Phys.*, 101: 819–841, 2000.

M. E. J. Newman. Clustering and preferential attachment in growing networks. *Phys. Rev. E*, 64: 025102, 2001a.

M. E. J. Newman. Fast algorithm for detecting communtity structure in networks. *Phys. Rev. E*, 69: 066133, 2004.

M. E. J. Newman. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci.*, 98: 404–409, 2001b.

M. E. J. Newman. Mixing patterns in networks. *Phys. Rev. E*, 67: 026126, 2003a.

M. E. J. Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89: 208701, 2002.

M. E. J. Newman. Scientific collaboration networks. i. network construction and fundamental results. *Phys. Rev. E*, 64: 016131, 2001c.

M. E. J. Newman. Scientific collaboration networks. ii. shortets paths, weighted networks, and centrality. *Phys. Rev. E*, 64: 016132, 2001d.

M. E. J. Newman. The structure and function of complex networks. *SIAM Review, Society for Industrial and Applied Mathematics*, 45: 167–256, 2003b.

M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69: 026113, 2004.

M. E. J. Newman and D. J. Watts. Renormalization group analysis of the small-world model. *Phys. Lett. A*, 263: 341–346, 1999a.

M. E. J. Newman and D. J. Watts. Scaling and percolation in the small-world network model. *Phys. Rev. E*, 60: 7332–7342, 1999b.

M. E. J. Newman and R. M. Ziff. Fast monte carlo algorithm for site or bond percolation. *Phys. Rev. E*, 64: 016706, 2001.

M. E. J. Newman, C. Moore, and D. J. Watts. Mean-field solution of the small-world network model. *Phys. Rev. Lett.*, 84: 3201–3204, 2000.

M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64: 026118, 2001.

M. E. J. Newman, A.-L. Barabási, and D. J. Watts. *The Structure and Dynamics of Networks*. Princeton University Press, Princeton, NJ, 2003.

NLANR. The national laboratory for applied network research. http://moat.nlanr.net/, www.

T. Odagaki and S Toyofuku. Properties of percolation clusters in a model granular system in two dimension. *J. Phys.:Condens. Matter*, 10: 6447–6452, 1998.

R. Olinky and L. Stone. Unexpected epidemic thresholds in heterogeneous networks: The role of disease transmision. *Phys. Rev. E*, 70: 030902(R), 2004.

S. A. Pandit and R. E. Amritkar. Charaterization and control of small-world networks. *Phys. Rev. E*, 60: R1119–R1122, 1999.

R. Pastor-Satorras and J. Rubi, editors. *Proceedings of the 18th Sitges Conference on Statistical Mechanics*. Lecture Notes in Physics. Springer-Verlag, Berlin, 2003.

R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86: 3200–3203, 2001.

R. Pastor-Satorras and A. Vespignani. *Evolution and Structure of the Internet*. Cambridge University Press, Cambridge, 2004.

R. Pastor-Satorras, A. Vázquez, and A. Vespignani. Dynamical and correlation properties of the internet. *Phys. Rev. Lett.*, 87: 258701, 2001.

A. Ramezanpour and V. Karimipour. Generating correlated networks from uncorrelated ones. *Phys. Rev. E*, 67: 046107, 2003.

A. Rapoport and W. J. Horvath. A study of a large sociogram. *Behavioral Sci.*, 6: 279–291, 1961.

T. S. Ray and N. Jan. Anomalous approach to the self-organized critical state in a model for "life at the edge of chaos". *Phys. Rev. Lett.*, 72: 4045–4048, 1994.

S. Redner. How popular is your paper? an empirical study of citation distribution. *Eur. Phys. J. B*, 4: 131, 1998.

H. E. Roman. A continuum percolation model for dispersed ionic conductors. *J. Phys.:Condens. Matter*, 2: 3909–3917, 1990.

A. F. Rozenfeld, R. Cohen, D. ben Avraham, and S. Havlin. Scale-free networks on lattices. *Phys. Rev. Lett.*, 89: 218701, 2002.

M. Sahimi. Long-range correlated percolation and flow and transport in heterogeneous porous media. *J. Phys. I*, 4: 1263–1268, 1994.

L. M. Sander, C. P. Warren, I. M. Sokolov, C. Simon, and J. Koopman. Percolation on heterogeneous networks as a model for epidemics. *Math. Biosci.*, 180: 293, 2002.

L. M. Sander, C. P. Warren, and I. M. Sokolov. Epidemics, disorder, and percolation. *Physica A*, 325: 1, 2003.

N. Schwartz, R. Cohen, D. ben Avraham, A.-L. Barabási, and S. Havlin. Percolation in directed scale-free networks. *Phys. Rev. E*, 66: 015104, 2002.

P. Sen. Accelerated growth in outgoing links in evolving networks: Deterministic versus stochastic picture. *Phys. Rev. E*, 69: 046107, 2004.

P. Sen and S. S. Manna. Clustering properties of a generalized critical euclidean network. *Phys. Rev. E*, 68: 026104, 2003.

P. Sen, S. Dasgupta, A. Chatterjee, P. A. Sreeram, G. Mukherjee, and S. S. Manna. Small-world properties of the indian railway network. *Phys. Rev. E*, 67: 036106, 2003.

V. D. P. Servedio, G. Caldarelli, and P. Buttà. Vertex intrinsic fitness: How to produce arbitrary scale-free networks. *Phys. Rev. E*, 70: 056126, 2004.

V. K. S. Shante and S. Kirkpatrick. An introduction to percolation theory. *Adv. Phys.*, 20: 235, 1971.

H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42: 425–440, 1955.

S. N. Soffer and A. Vázquez. Network clustering coefficient without degree-correlation biases. *Phys. Rev. E*, 71: 057101, 2005.

I. M. Sokolov. Dimensionalities and the other geometric critical exponents in percolation theory. *Sov. Phys. Usp.*, 29: 924, 1986.

R. Solomonoff and A Rapoport. Connectivity of random nets. *Bull. Math. Biophysics*, 13: 107–117, 1951.

O. Sotolongo-Costa and G. J. Rodgers. Bose-einstein condensation in random directed networks. *Phys. Rev. E*, 68: 056118, 2003.

H. E. Stanley. *Introduction to Phase Transitions and Critical Phenomena*. The international series of monographs on physics. Oxford University Press, 1971.

D. Stauffer. Scaling theory of percolation clusters. *Phys. Rep.*, 54: 1–74, 1979.

D. Stauffer and A. Aharony. *Introduction to Percolation Theory*. Taylor and Francis, 2nd edition edition, 1994.

J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E. D. Gilles. Metabolic network structure determines key aspects fo funtionality and regulation. *Nature*, 420: 190–193, 2002.

S. H. Strogatz. Exploring complex networks. *Nature*, 410: 268–276, 2001.

SW. Small world project. http://smallworld.columbia.edu/, www.

B. Tadić. Dynamics of directed graphs: The world-wide web. *Physica A*, 293: 273–284, 2001.

B. Tadić. Temporal fractal structures: Origin of power laws in the world-wide web. *Physica A*, 314: 278–283, 2002.

C. Tsallis and M.P. de Albuquerque. Are citations of scientific papers a case of nonextensivity? *Euro. Phys. J. B*, 13: 777, 2000.

P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, and J. M. Rothberg. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*, 403: 623–627, 2000.

A. Vázquez. Statistics of citation networks. *Preprint 0105031*, 2001.

A. Vázquez and Y. Moreno. Resilence to damage of graphs with degree correlations. *Phys. Rev. E.*, 67: 015101, 2003.

A. Vázquez, R. Pastor-Satorras, and A. Vespignani. Large-scale topological and dynamical properties of internet. *Phys. Rev. E.*, 65: 066130, 2002.

A. Vázquez, A. Flammini, A. Maritan, and A. Vespignani. Modeling of protein interaction networks. *Complexus*, 1: 38–44, 2003.

D. Volchenkov, L. Volchenkova, and Ph. Blanchard. Epidemic spreading in a variety of scale free networks. *Phys. Rev. E*, 66: 046137, 2002.

E. Volz. Random networks with tunable degree distribution and clustering. *Phys. Rev. E*, 70: 056115, 2004.

A. Wagner. How the global structure of protein interaction networs evolves. *Proc. R. Soc. Lond. B*, 270: 457–466, 2003.

A. Wagner and D. Fell. The small world inside large metabolic networks. *Technical Report No. 00-07-041, Santa Fe Institute*, 2000.

X. F. Wang and G. Chen. Complex networks: Small-world, scale-free and beyond. *IEEE Circuits and Systems Magazine*, 3: 6–20, 2003.

C. P. Warren, L. M. Sander, and I. M. Sokolov. Geography in a scale-free network model. *Phys. Rev. E*, 66: 056105, 2002.

D. J. Watts. *Small Worlds*. Princeton University Press, NJ, 1999a.

D. J. Watts. Networks, dynamics, and the small world phenomenon. *Amer. J. Sociol.*, 105: 493–592, 1999b.

D. J. Watts. *Six Degrees: The Science of a Connected Age*. Norton, New York, 2003.

D. J. Watts and S. H. Strogatz. Collective dynamics of "small-world" networks. *Nature*, 393: 440–442, 1998.

J. C. Wierman. Percolation theory. *Ann. Prob.*, 10: 509–524, 1982.

K. G. Wilson. The renormalization group and critical phenomena. *Rev. Mod. Phys.*, 55: 583, 1982.

K. G. Wilson and J. Kogut. An early exposition of the renormalization group in quantitative form. *Phys. Rep.*, 12C: 75, 1974.

R. Xulvi-Brunet and I. M. Sokolov. Evolving networks with disadvantaged long-range connections. *Phys. Rev. E*, 66: 026118, 2002.

R. Xulvi-Brunet and I. M. Sokolov. Reshuffling scale-free networks: From random to assortative. *Phys. Rev. E*, 70: 066102, 2004.

R. Xulvi-Brunet and I. M. Sokolov. Changing correlations in networks: assortativity and dissortativity. *Acta Physica Polonica B*, 36: 1431–1455, 2005.

R. Xulvi-Brunet, W. Pietsch, and I. M. Sokolov. Correlations in scale-free networks: Tomography and percolation. *Phys. Rev. E*, 68: 036119, 2003.

K. Yang, L. Huang, and L. Yang. Lattice scale-free networks with weighted linking. *Phys. Rev. E*, 70: 015102, 2004.

YEAST. *Saccharomyces* genome database. www.yeastgenome.org, www.

S.-H. Yook, H. Jeong, and A.-L. Barabási. Modelling the internet's large-scale topology. *Proc. Natl. Acad. Sci. USA*, 99: 13382–13386, 2002.

R. Zallen. *The physics of Amorphous Solids.* Wiley, New York, 1983.

D. H. Zanette. Dynamics of rumor propagation on small-world networks. *Phys. Rev. E*, 65: 041908, 2002.

D. Zheng, S. Trimper, B. Zheng, and P. M. Hui. Weighted scale-free networks with stochastic weight assignments. *Phys. Rev. E*, 67: 040102, 2003.

H. Zhu, X. Wang, and J.-Y. Zhu. Effect of aging on network structure. *Phys. Rev. E*, 68: 056121, 2003.

Acknowledgments

I thank my Ph.D. supervisor Prof. Dr. Igor Sokolov for his invaluable help, support, guidance, and patience during the research leading to this thesis. It would have been impossible to realize this work without his advice and suggestions. I would like also to acknowledge numerous enlightening discussions about the work and noteworthy improvements of this text to Wolfgang Pietsch. Finally, I am indebted to Eleanor Livesey, Wolfgang Pietsch, Cristina Sánchez Pelegrín, Luis Eduardo Ortiz Tánchez, and Jesús Emeterio Navarro Barrientos for their help with the English language, who read the manuscript and corrected various mistakes.

Anyway, aquesta tèsi no l'havera pogut escriure sense les persones que han fet de la meua estància a Berlí una experiència única i irrepetible: els amics. La llista és llarga. Per a començar, m'agradaria citar a tots aquells amb els que ja no tic massa contacte pero que encara recorde a sovint: Kirsten Ziegenmeier, Zora, Norikazo, Eliana, Lupe, María, Oscar, Samuel, Roberto, Francesco, Natalia, Anne, Gabriela, Katja, Kim, Milena, Ulrike, Marina, Ingo, Steffi, David, Stephani, Susanne, Thoralf, Silvia, Sven, Idrees, Sophie, Imat, Mercedes, Manish, Abe, Juan, Magda, Åsa, Wendy, Ingrid, Lucía, Rahel, etc. Sehr wichtig waren und sind meine Kollegen am Institut für Physik Wolfgang Pietsch, Robert Engel, Eduardo Ortiz, Emeterio Navarro, Tania Verechtchagina, Jana Wehrmeister, Thilo Schwalger, Xaver Sailer und Tobias Prager. Weitere sehr gute Freunde in Berlin sind: Martin Obermayr, Kristin Moers, Juan Barres, Oscar Fuster, Emilio Rincón, Teresa Palomar, Jutta Ortiz, Gunnar Lorezen, Ana Carolina Garcia, Katharina Meier und Beate Freundlich. Besondere Erwähnung verdient Claire Kremer, mit der ich mich verstehen kann, ohne Wörter zu sagen. Termino con aquellos que han conseguido sacarme de casa incluso en estos últimos meses en los que me he dedicado casi exclusivamente a escribir la tesis (sobra decir que tipo de amigos son...): Cristina Sánchez, Alfonso Gutiérrez, Sergio Alonso, Emeterio Navarro, Eduardo Ortiz y Leticia Díaz Beltrán. A ésta última, jiennense más chula que un ocho, va además dedicado el apartado de "agradables desagradecimientos", pues su sola presencia por Berlín hizo que olvidara durante demasiados meses ésta tesis y demás proyectos académicos...

Selbständigkeitserklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig ohne fremde Hilfe verfaßt und nur die angegebene Literatur und Hilfsmittel verwendet zu haben.

Ramon Xulvi-Brunet, Berlin, den 28.4.2006.